

BIOLOGIA MOLECOLARE



DALLA SCOPERTA DEL DNA AL CODICE GENETICO E STRUTTURA DEGLI ACIDI NUCLEICI

Questa prima parte del libro introduce la Biologia Molecolare in modo storico, mostrando come nella prima metà del secolo scorso sia stata fatta luce su una caratteristica della vita che sembrava incomprensibile, cioè come avviene la trasmissione dei caratteri ereditari. Infatti, tutti gli organismi viventi del nostro pianeta hanno bisogno, per portare avanti i loro cicli vitali, delle istruzioni contenute nella struttura degli acidi nucleici, DNA e RNA. Queste informazioni, che possono essere paragonate al software di un computer, utilizzano un linguaggio, il codice genetico, che è risultato essere sostanzialmente universale, cioè uguale per tutti gli esseri viventi.

Descriveremo quindi gli eventi che hanno portato alla scoperta del DNA come materiale genetico, alla risoluzione della sua struttura a doppia elica e alla decifrazione del codice genetico. In particolare, descriveremo il background culturale che ha portato a questi importanti risultati e, soprattutto, il fatto che è stata l'unione di molte discipline diverse a produrre quel cambiamento di mentalità che ha permesso di affrontare e risolvere, in modo del tutto nuovo, il problema dell'identificazione e della struttura del materiale ereditario, fornendo le basi molecolari alla teoria evoluzionistica di Darwin. Spiegare la struttura dei geni e la loro espressione come il frutto d'interazioni tra molecole ha cambiato per sempre la nostra visione della Biologia.

1

Introduzione alla Biologia Molecolare

1.1 Che cos'è la Biologia Molecolare?

1.2 Il gruppo del fago e la nascita della Biologia Molecolare

1.3 Dalla scoperta del DNA alla dimostrazione del suo ruolo come materiale genetico

1.1 Che cos'è la Biologia Molecolare?

Cosa sia la Biologia Molecolare e quali siano le sue relazioni con le altre discipline biologiche è stato a lungo dibattuto, e lo è tuttora prestandosi l'argomento a diversi punti di vista. L'espressione "Biologia Molecolare" potrebbe essere estesa a tutti gli studi a livello di molecole dei sistemi viventi, quindi a quasi tutta la Biologia, tranne l'Ecologia e gli aspetti naturalistici di Zoologia, Botanica e Microbiologia. D'altra parte oggi tutto il resto è arrivato al livello molecolare, la Genetica, la Fisiologia, la Biologia dello Sviluppo, la Virologia, l'Immunologia, per non nominare la Biochimica, che al livello molecolare c'è sempre stata.

La definizione della disciplina Biologia Molecolare non può, quindi, fare riferimento al livello molecolare di studio ma, con un po' di semplificazione, è possibile intenderla, nel senso storico del progredire delle conoscenze, come un'area scientifica che a partire dagli anni quaranta del secolo scorso ha permesso di unificare due aree che erano state sempre separate, cioè la Genetica e la Biochimica.

La Biochimica la troviamo già come importante area di ricerca all'inizio del Novecento e, come la Fisiologia, era praticata essenzialmente in ambienti medici che studiavano l'organismo, soprattutto umano, per comprenderne il funzionamento a livello chimico. Campi d'interesse della Biochimica erano il metabolismo, la catalisi, gli enzimi e in genere la struttura e funzione delle proteine.

Negli stessi anni si comincia a sviluppare la Genetica. Nata con il lavoro di Gregory Mendel nella seconda metà del XIX secolo, essa si sviluppa e diffonde agli inizi del Novecento in un ambiente culturale molto diverso dalla Biochimica. Infatti, mentre quest'ultima si sviluppava nell'ambiente delle ricerche mediche, la Genetica voleva soprattutto rispondere all'esigenza di agricoltori e allevatori di comprendere i meccanismi della trasmissione dei caratteri ereditari al fine di poter migliorare le specie animali e vegetali d'interesse agronomico e trovava presto riscontro in un ambiente culturale biologico.

Oltre a svilupparsi in differenti ambienti culturali, la Biochimica e la Genetica usavano approcci scientifici molto diversi. La Biochimica, seguendo un approccio tipicamente "riduzionistico", isolava, purificava e analizzava i vari componenti molecolari degli organismi viventi, con l'idea che capire come funzionano le parti di un sistema complesso è essenziale per capire come funziona l'insieme. La Genetica, invece, utilizzava l'approccio opposto, detto "olistico", cioè affrontava un

Tabella 1.1 Cronologia dei risultati della Genetica, della Biochimica e della Biologia Molecolare sino alla fine degli anni cinquanta.

Genetica		Biochimica
1865	I geni sono dei fattori particolati	1879 scoperta del DNA
1900	I cromosomi sono unità ereditarie	Analisi chimica e studi biochimici del DNA
1910	I geni sono localizzati nei cromosomi	
1930	Le mutazioni sono cambiamenti fisici nei geni	
Biologia Molecolare		
1944	Il DNA è il materiale genetico	
1953	Il DNA ha una struttura a doppia elica	
1959	Il DNA si replica in modo semiconservativo	

fenomeno complesso, come la trasmissione dei caratteri ereditari, studiandone gli aspetti macroscopici a livello fenotipico per ottenere informazioni sui livelli sottostanti come l'organizzazione del materiale genetico. Queste differenze metodologiche implicavano anche l'uso di sistemi biologici diversi: mentre la Genetica, che era iniziata con i piselli di Mendel, aveva avuto grandi sviluppi con altri organismi, soprattutto *Drosophila*, ma anche *Neurospora* e lievito, mais ecc., la Biochimica utilizzava soprattutto organi e tessuti di piante e animali superiori (per es. fegato di ratto, di maiale ecc.) per ottenere massicce preparazioni delle molecole che voleva analizzare. Queste differenze di origine culturale, di sistemi biologici e di approcci sperimentali utilizzati, facevano sì che le due aree di ricerche avessero pochi contatti, anzi vi erano spesso diffidenze, incomprensioni e rivalità.

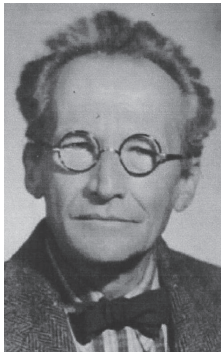
È così che negli anni trenta e quaranta del secolo scorso, nonostante gli importanti sviluppi e successi sia della Biochimica che della Genetica, restava un grande divario tra le due aree. Sono gli anni in cui, sentendo l'esigenza di una maggiore unitarietà nello studio dei sistemi viventi e con l'apporto di nuovi approcci culturali e metodologici, ha avuto inizio e si è sviluppata la nuova area della Biologia Molecolare che, quindi, definiremmo storicamente come quell'area di ricerca che ha avuto come primo grande successo quello di spiegare il fenomeno della trasmissione dei caratteri ereditari in termini molecolari, operando così un'unificazione della Genetica e della Biochimica. Oggi la Biologia Molecolare ha invaso, con i suoi concetti e i suoi metodi sperimentali, quasi tutti i campi della Biologia, e non solo (tab. 1.1).

1.2 Il gruppo del fago e la nascita della Biologia Molecolare

A questi nuovi sviluppi hanno contribuito ambedue gli ambienti culturali della Biochimica e della Genetica, ma l'impulso decisivo è venuto dall'esterno, e in particolare da un gruppo di fisici, perlopiù tedeschi, che in Germania gravitavano intorno a Niels Bohr e alla sua scuola, i quali rivolsero la loro attenzione ai fondamentali problemi della Biologia. Ricordiamo tra questi Erwin Schrödinger (1887-1961), uno dei padri della meccanica quantistica che, nel 1944, pubblicò un breve libro intitolato *Che cos'è la vita?* in cui veniva posto con chiarezza il problema (ma non la soluzione) della struttura dei geni (fig. 1.1). Tra questi scienziati, il personaggio di maggior spicco è stato il fisico Max Delbrück (1906-1981) che, nel 1937, lasciò la Germania nazista e si trasferì negli Stati Uniti dove si dedicò totalmente alla Biologia. Insieme a Salvador Luria (1912-1991), italiano laureato in Medicina che si era trasferito negli Stati Uniti a causa delle leggi razziali promulgate in Italia dal regime fascista, costituì quello che fu

chiamato negli anni successivi il **gruppo del fago**, un gruppo di ricercatori di varia origine (fisici, medici, biologi) che scelsero di utilizzare i batteriofagi per studiare il problema della struttura e funzione dei geni (**fig. 1.2**). L'idea era che per studiare un fenomeno complesso come la trasmissione dei caratteri ereditari e, quindi, la struttura e funzione dei geni, bisognasse cominciare con lo studiare un sistema semplificato. Lasciarono perdere quindi piselli e *Drosophile* e scelsero di studiare i batteri e soprattutto i batteriofagi, che sono i virus che infettano i batteri (**fig. 1.3**). Questi, detti comunemente fagi, erano considerati come i

A



B

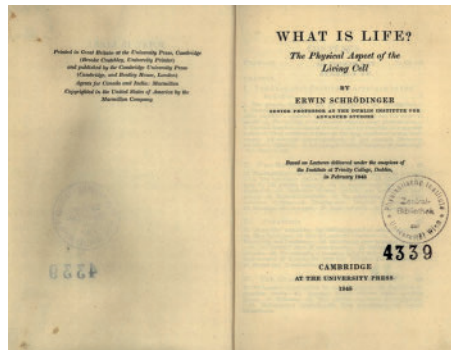
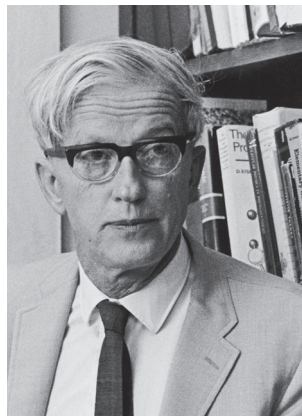


Figura 1.1 Erwin Schrödinger e il suo libro *Che cos'è la vita?*. (A) Il fisico austriaco Erwin R.J.A. Schrödinger (1887-1961) è stato uno dei grandi scienziati del XX secolo e uno dei padri fondatori della Meccanica quantistica. (B) Frontespizio del suo libro *What is life?* pubblicato nel 1944 e basato su una serie di tre lezioni pubbliche svolte al Trinity College di Dublino, nel febbraio del 1943.

A

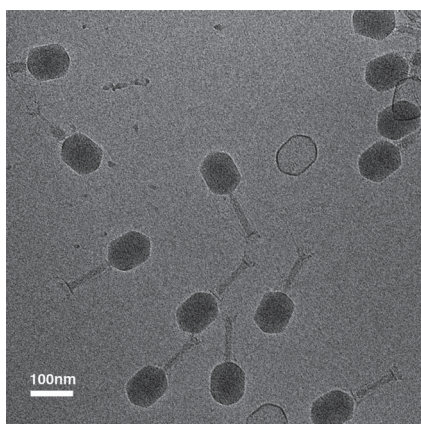


B

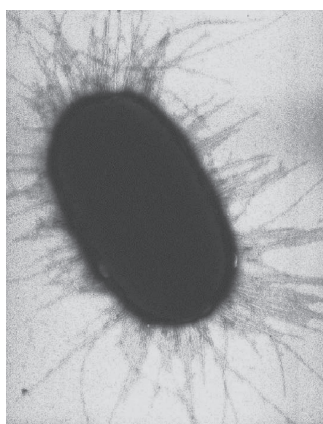


Figura 1.2 I fondatori e iniziatori del "gruppo del fago". (A) Max Delbrück (1906-1981). (B) Salvador Luria (1912-1991).

A



B



C

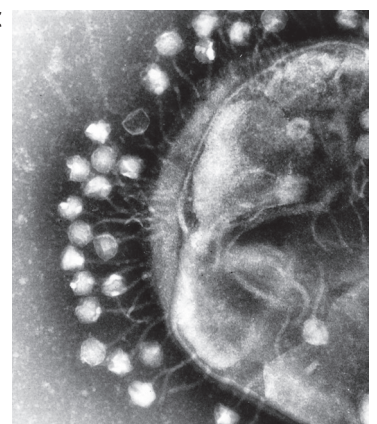


Figura 1.3 Il fago T4 e il suo ospite *Escherichia coli*. Immagini al microscopio elettronico di (A) particelle fagiche mature, (B) batterio *Escherichia coli*, (C) particelle fagiche mentre infettano una cellula ospite.

sistemi genetici più semplici e si pensava che potessero contenere uno o pochissimi geni. Oggi sappiamo che, in realtà, anche se esistono piccoli fagi contenenti pochi geni, in genere sono più complessi, contenendone anche cinquanta, cento o di più. Negli anni quaranta e nei primi anni cinquanta del secolo scorso il gruppo del fago mise a punto metodi, sistemi e concetti che saranno alla base della nuova Biologia Molecolare.

1.3 Dalla scoperta del DNA alla dimostrazione del suo ruolo come materiale genetico

Anche se attirerà l'interesse dei biologi solo negli anni 1940-1950, il DNA era stato scoperto molto prima. Infatti, intorno al 1870, il chimico svizzero Friedrich Miescher (1844-1895), che lavorava presso l'Università di Tubinga in Germania, isolò dai globuli bianchi del sangue una sostanza ricca di fosforo che chiamò "nucleina", di cui cominciò a studiare le proprietà chimiche portando così alla scoperta degli acidi nucleici e, quindi, del DNA. È curioso notare che Miescher lavorava a poche centinaia di chilometri da Brno, dove negli stessi anni Gregory Mendel stava compiendo i suoi studi sui geni. Ma questi due concetti, gene e DNA, sarebbero rimasti separati per oltre 70 anni!

Nei decenni successivi alla scoperta di Miescher, la **nucleina** (poi acido timonucleico, acido nucleico e infine acido deossiribonucleico = DNA) divenne oggetto di vari studi volti a caratterizzarne la struttura chimica. Fu subito chiarito che, come sarà descritto in dettaglio in seguito, il DNA è un polimero di nucleotidi in cui si alternano uno zucchero e un gruppo fosfato, mentre lateralmente alla catena sono legate le basi azotate, che nel DNA sono di quattro tipi: adenina, citosina, guanina e timina. Per quanto riguarda il peso molecolare del DNA (e quindi la lunghezza della catena in termini di numero di nucleotidi) oggi sappiamo che è molto grande, ma che durante le procedure di purificazione ed estrazione dalle cellule può subire una progressiva degradazione a molecole sempre più piccole, soprattutto se, come succedeva a quell'epoca, non si usano particolari precauzioni e accorgimenti. È così che le stime di peso molecolare del DNA ottenute a partire dalla scoperta di Miescher, e ripetute altre volte per oltre quarant'anni successivi, davano un peso molecolare approssimativo di poco più di 1000 Da. Il caso vuole che questo peso molecolare corrisponda approssimativamente a quello di 4 nucleotidi. Poiché i tipi di basi che costituiscono i nucleotidi sono 4 (adenina, citosina, guanina e timina) e queste sembravano essere presenti approssimativamente in quantità uguali, fu proposto il **modello del tetranucleotide**, che descriveva il DNA come una molecola piccola, costituita da soli quattro nucleotidi con una ciascuna delle quattro basi (**fig. 1.4**). È chiaro, quindi, perché ancora dopo tanti anni dalla sua scoperta il DNA non venisse preso in considerazione per alcun ruolo sofisticato nei processi cellulari, tanto meno per quello di gene.

La **figura 1.5** mostra come a partire dagli anni trenta le misure di peso molecolare del DNA cominciarono a dare risultati diversi. Infatti, le tecniche di preparazione e manipolazione del DNA diventarono, col passare degli anni, sempre più raffinate e mostrarono che il DNA ha sempre un peso molecolare molto alto, in certi casi eccezionalmente alto. Questo cambiamento di nozione di peso molecolare è stato un passo essenziale verso la scoperta del ruolo del DNA come materiale genetico.

Nel frattempo, altre importanti ricerche per l'identificazione del DNA come materiale genetico vennero fatte in un ambiente scientifico diverso. Il medico e microbiologo inglese Frederick Griffith (1879-1941) stava studiando la possibilità di fare un vaccino per prevenire la polmonite e per i suoi studi utilizzava due ceppi del batterio *Streptococcus pneumoniae*: il ceppo R e il ceppo S (R e S

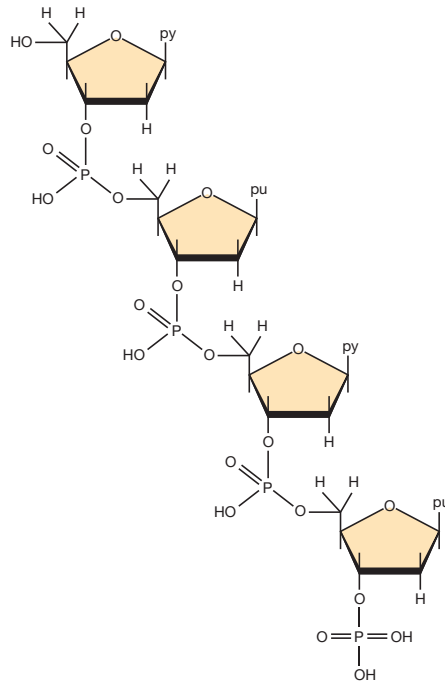
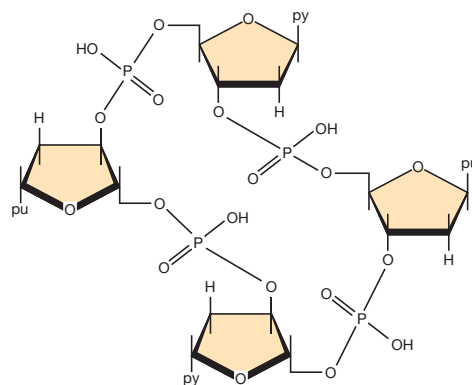
A**B**

Figura 1.4 Il tetranucleotide. Il modello del tetranucleotide proposto da Phoebus A. Levene e comunemente accettato nei primi decenni del secolo scorso, descrive il DNA come piccole molecole costituite ciascuna da quattro nucleotidi, escludendo così la possibilità che il DNA possa costituire il materiale genetico. **(A)** Modello di tetranucleotide lineare. **(B)** Modello di tetranucleotide ciclico.

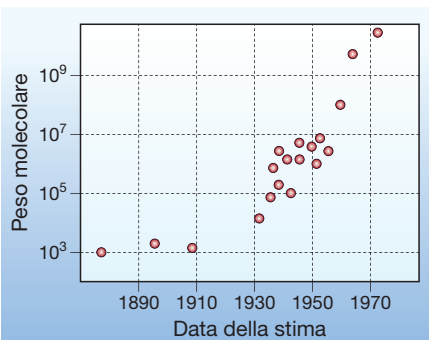
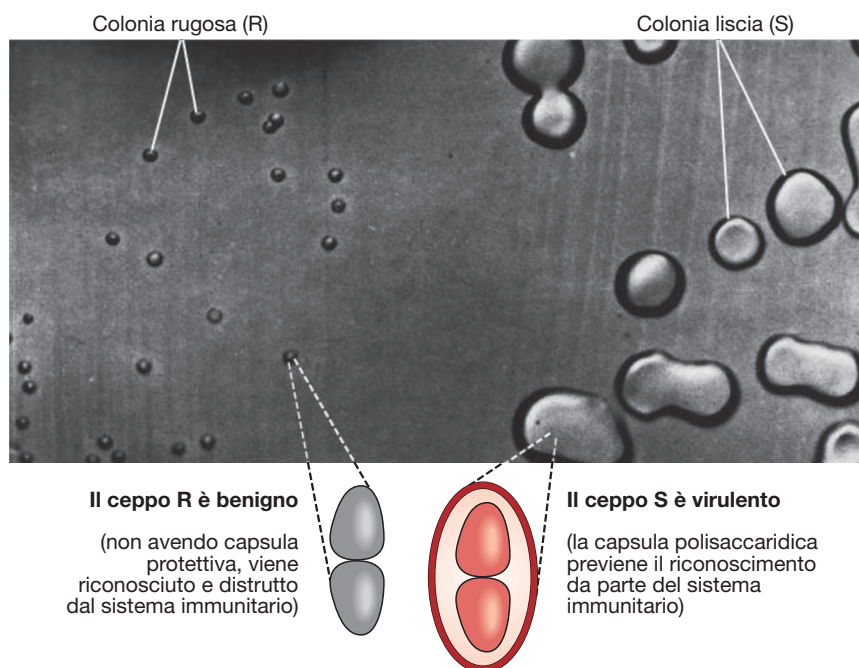


Figura 1.5 Peso molecolare del DNA. La stima del peso molecolare del DNA è andata aumentando nel corso degli anni, con il raffinamento delle tecniche di estrazione e purificazione.

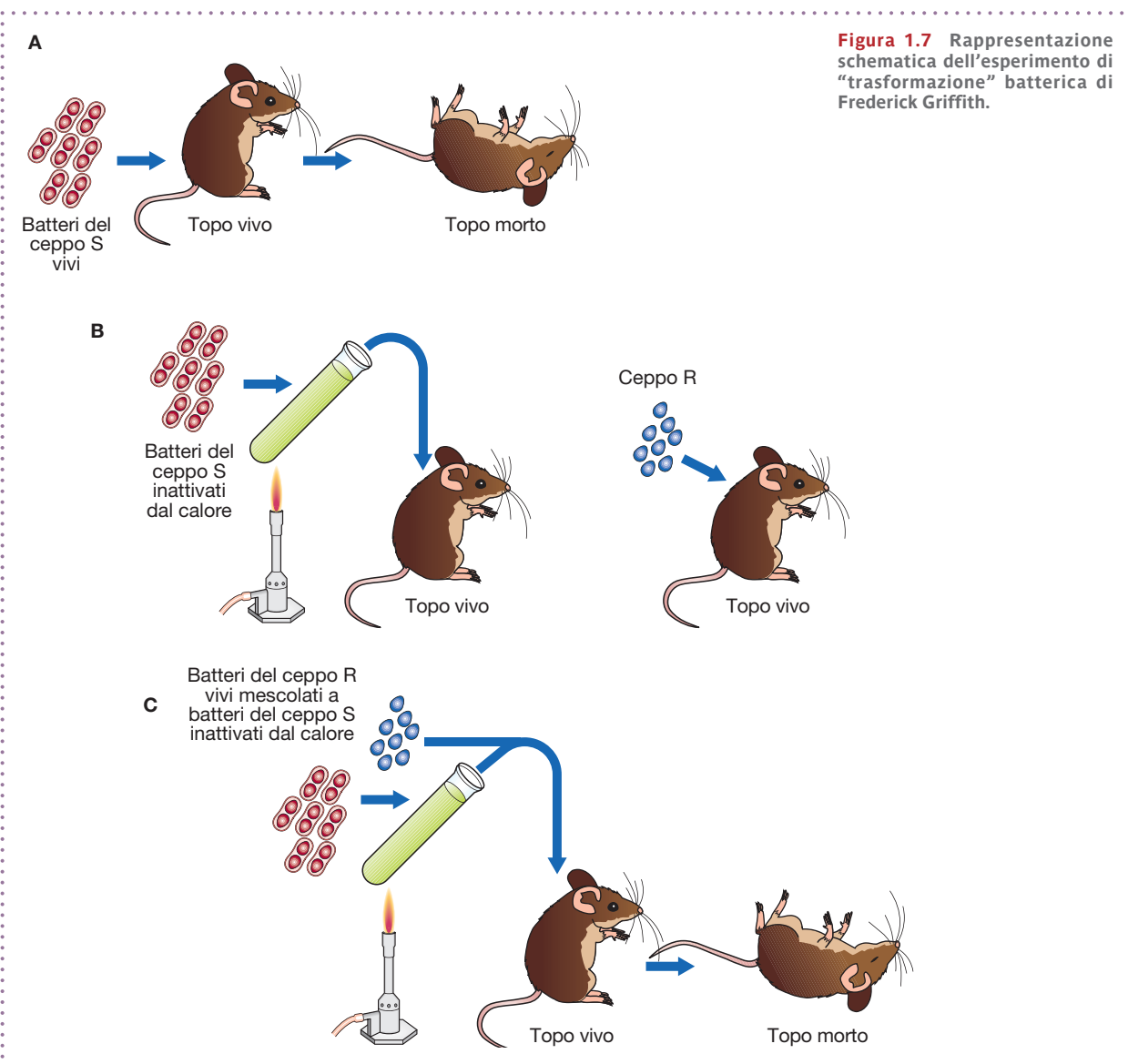
stanno per *rough* e *smooth*, cioè “rugoso” e “liscio”, a seconda dell'apparenza della colonia di batteri cresciuta in una piastra di agar (**fig. 1.6**). La **figura 1.7** mostra lo schema dell'importante esperimento con cui Griffith scoprì, nel 1928, il principio trasformante. Il ceppo R, che è privo di una capsula polisaccaridica,

Figura 1.6 *Streptococcus pneumoniae*: il ceppo benigno R (*rough*, rugoso) e il ceppo virulento S (*smooth*, liscio). Si noti come le colonie lisce sono più grandi a causa della capsula polisaccaridica che avvolge ciascun batterio.



non causa polmonite nei topi in cui viene iniettato perché questi si difendono con il loro sistema immunitario. Al contrario, il ceppo S, che ha una capsula polisaccaridica che lo protegge dalle difese immunitarie dell'animale, causa polmonite e successiva morte nei topi in cui viene iniettato. Quando il ceppo S viene inattivato (ucciso) alzando la temperatura e poi iniettato nei topi, perde l'effetto patogeno. Tuttavia, quando il ceppo S inattivato viene mescolato col ceppo R e i due sono iniettati insieme nei topi, questi contraggono polmonite e muoiono. Isolando i batteri dal sangue di questi topi morti, Griffith scoprì che il ceppo R normalmente non patogeno acquisisce la capsula polisaccaridica e mantiene questa caratteristica fenotipica per molte generazioni. Griffith ipotizzò che un qualcosa del ceppo S inattivato penetrasse nel ceppo R convertendolo in un S virulento. Lui chiamò questo qualcosa **principio trasformante** (oggi diremmo gene/i per la produzione della capsula polisaccaridica).

Questo esperimento aprì la strada al successivo passo fondamentale fatto, presso il Rockefeller Institute Hospital a New York, da Oswald T. Avery (1877-1955) (**fig. 1.8A**) e dai suoi collaboratori Colin MacLeod e Maclyn McCarty. Come descrivono questi autori nel loro lavoro pubblicato nel 1944, essi ripresero l'esperimento di Griffith e, con le tecniche di frazionamento che andavano affinandosi, affrontarono l'identificazione, nell'estratto cellulare del ceppo S, del componente chimico che ha la proprietà di principio trasformante. Una semplificazione sperimentale che Avery e collaboratori introdussero rispetto all'esperimento di Griffith fu che, anziché iniettare i vari ceppi di pneumococco nei topi per verificare la loro patogenicità, questi venivano identificati come S o R semplicemente osservando la superficie (ruvida o liscia) delle colonie su piastra di agar. La domanda era: quale dei componenti macromolecolari presenti nell'estratto di batteri di ceppo S (proteine, lipidi, polisaccaridi, DNA, RNA) è l'effettivo responsabile della trasformazione? Avery e collaboratori fecero degli estratti dai batteri patogeni che distribuirono in varie provette per dimostrare la natura del principio trasformante. Ciascuna provetta conteneva un diverso reagente per la degradazione specifica di un componente molecolare: tripsina che digerisce le proteine, RNasi che degrada l'RNA, DNasi che degrada il DNA ecc. A tutte



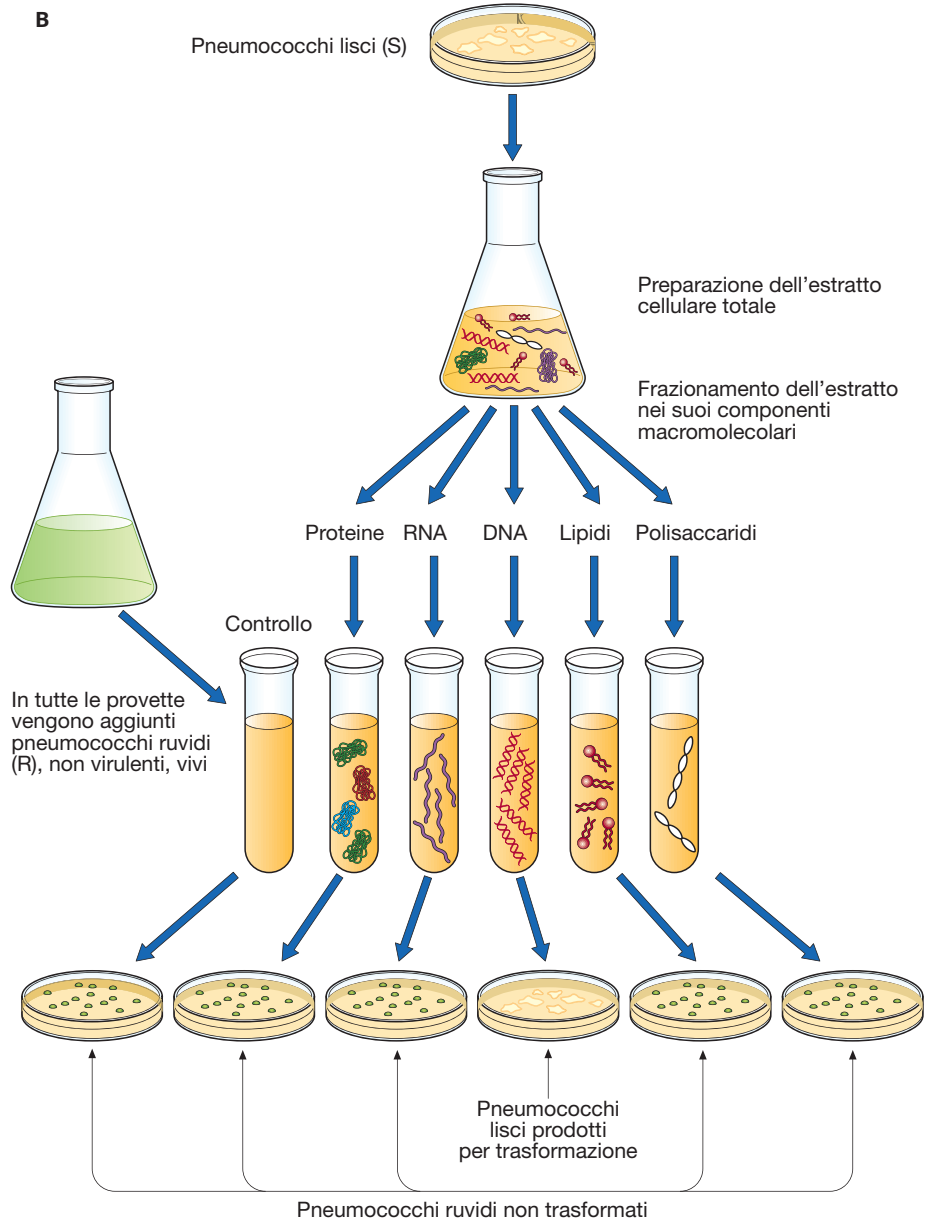
le provette venivano aggiunti i batteri rugosi non patogeni che dovevano essere trasformati. Solo l'eliminazione del DNA impediva la trasformazione dei batteri da rugosi a lisci e patogeni, dimostrando che era questa molecola il principio trasformatore. Avery e colleghi fecero inoltre delle vere e proprie purificazioni delle componenti degli estratti, usando negli esperimenti proteine, RNA e DNA purificati, come illustrato in **figura 1.8B**.

Questa scoperta, che per molti anni è stata chiamata la **bomba di Avery** per indicare l'enorme impatto che successivamente ebbe nei progressi dello studio del materiale genetico, arrivò in un certo senso prematura, quando ancora gran parte dell'ambiente scientifico non era in grado di accettarla. Infatti, per vari motivi, anche i migliori ambienti scientifici avevano una forte propensione a considerare come materiale genetico le proteine, piuttosto che il DNA. Infatti, le proteine si stavano dimostrando molecole dalle straordinarie possibilità strutturali e funzionali, mentre il DNA veniva per lo più considerato una molecola troppo semplice. Si pensava che per codificare molecole complesse come le proteine anche i geni dovessero essere almeno altrettanto complessi, se non di più. Bisogna dire che anche



A

Figura 1.8 La “bomba di Avery”. (A) Oswald T. Avery (1877-1955). (B) Partendo da un ceppo di pneumococchi lisci e virulenti, Avery e collaboratori fecero un estratto cellulare che era in grado di trasformare pneumococchi del ceppo ruvido benigno in liscio e virulento. In seguito purificarono le diverse componenti molecolari dell’estratto (proteine, DNA, lipidi ecc.). Solo il DNA era in grado di operare la trasformazione, dimostrando che il DNA è il principio trasformante.



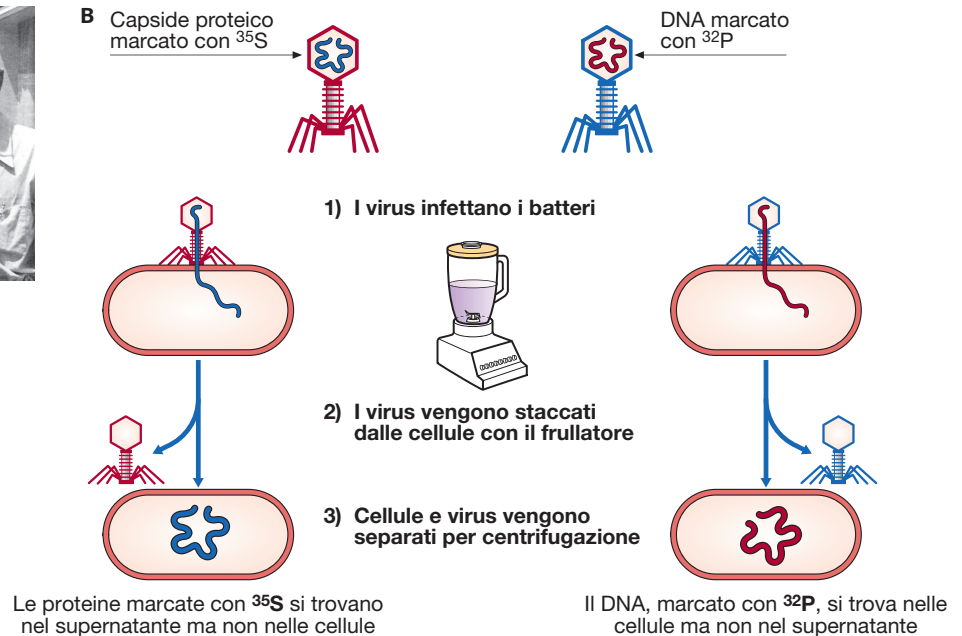
Delbrück e i vari ricercatori che si riconoscevano nel gruppo del fago puntavano in quegli anni alle proteine come base strutturale dei geni. Fatto sta che la scoperta di Avery e collaboratori ebbe poco credito per qualche anno, finché proprio dal gruppo del fago venne pubblicato nel 1952 un esperimento di tipo completamente diverso che confermò il DNA come materiale genetico. Questo esperimento, chiamato scherzosamente l'**esperimento del frullatore**, fu fatto da Alfred Hershey e Martha Chase (**fig. 1.9A**) nel Laboratorio di Cold Spring Harbor a Long Island, vicino a New York, utilizzando i batteriofagi.

Un fago è un piccolo virus che infetta i batteri e consiste di un capsido proteico che ne racchiude il materiale genetico. Quando un fago infetta un batterio, vi inserisce il materiale genetico mentre il capsido rimane al di fuori. Hershey e Chase marcarono radioattivamente il DNA del fago con fosforo radioattivo (^{32}P) e le proteine del capsido con zolfo radioattivo (^{35}S). Subito dopo l'infezione del batte-



A

Figura 1.9 (A) Martha Chase (1927-2003) e Alfred Hershey (1908-1997) al Cold Spring Harbor Laboratory nel 1953. (B) L'esperimento del "frullatore" di Chase e Hershey. Le molecole marcate con isotopi radioattivi sono disegnate in rosso.



rio con i fagi marcati nel brodo di coltura, questo veniva frullato con un normale frullatore, che provocava il distacco del capsido fagico vuoto dai batteri in cui era entrato il materiale genetico del fago. La successiva centrifugazione del brodo di coltura consentì di verificare che il DNA marcato con ^{32}P si trovava all'interno dei batteri, mentre le proteine marcate con ^{35}S erano rimaste fuori. Questo esperimento dimostrò che il DNA è il materiale genetico del fago e che le proteine non trasmettono informazione genetica (**fig. 1.9B**).

Va sottolineato che gli esperimenti di Avery erano di per sé più convincenti a dimostrare il ruolo del DNA, ma arrivarono in tempi prematuri. L'esperimento del frullatore, invece, trovò un ambiente scientifico pronto ad accogliere l'idea del DNA come materiale genetico e, soprattutto, aveva la forza convincente di venire dal gruppo del fago con approcci e tecniche nuove. Comunque sia, nei primi anni cinquanta praticamente tutto l'ambiente scientifico più importante accettava l'idea che il DNA è il vero materiale genetico. Cominciò da allora la corsa per comprendere la struttura fisica tridimensionale del DNA, con la certezza che questa avrebbe fatto luce sulla struttura e la funzione dei geni.

Lecture di approfondimento e siti web

Cairns, J., Stent, G.S. e Watson, J.D., *Phage and the origin of molecular biology*, Cold Spring Harbor Laboratory Press, NY, 1966.

Judson, H.F., *The eighth day of creation*, Cold Spring Harbor Laboratory Press, NY, 1996.

Edizione italiana: *L'ottavo giorno della creazione*, Editori Riuniti, Roma, 1986.

<http://www.dnalc.org/>

<http://www.dnai.org/>

<http://www.dnaftb.org/>

2

Struttura degli acidi nucleici

2.1 Struttura chimica degli acidi nucleici

2.2 Struttura fisica del DNA: la scoperta della struttura a doppia elica

2.3 Struttura fisica del DNA: i parametri strutturali della doppia elica

2.4 Topologia del DNA e DNA topoisomerasi

2.5 Struttura dell'RNA

2.1 Struttura chimica degli acidi nucleici

Gli acidi nucleici, DNA e RNA, sono costituiti da catene polinucleotidiche, cioè polimeri lineari di unità chiamate **nucleotidi**. Descriveremo prima i singoli nucleotidi isolati, per poi vedere come questi sono legati tra loro in una catena di acido nucleico.

I nucleotidi sono molecole costituite da tre componenti: uno zucchero pentoso, una base azotata, uno o più gruppi fosfato.

Nel caso del DNA lo zucchero è il **deossiribosio** (chiamato anche desossiribosio), mentre nell'RNA è il **ribosio**. Come si vede in **figura 2.1**, i due zuccheri

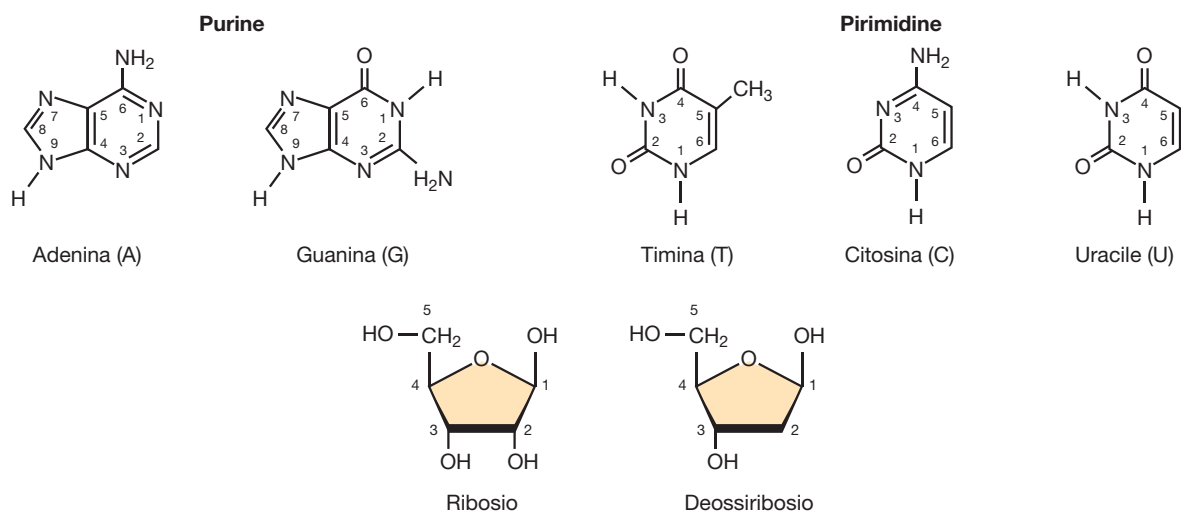


Figura 2.1 Strutture dei componenti chimici degli acidi nucleici.

Tabella 2.1 Basi, nucleosidi e nucleotidi.

Basi	Nucleosidi	Nucleotidi	Abbreviazione	
			RNA	DNA
Adenina	Adenosina	Acido adenilico	AMP	dAMP
Guanina	Guanosina	Acido guanilico	GMP	dGMP
Citosina	Citidina	Acido citidilico	CMP	dCMP
Timina	Timidina	Acido timidilico		dTMP
Uracile	Uridina	Acido uracilico	UMP	

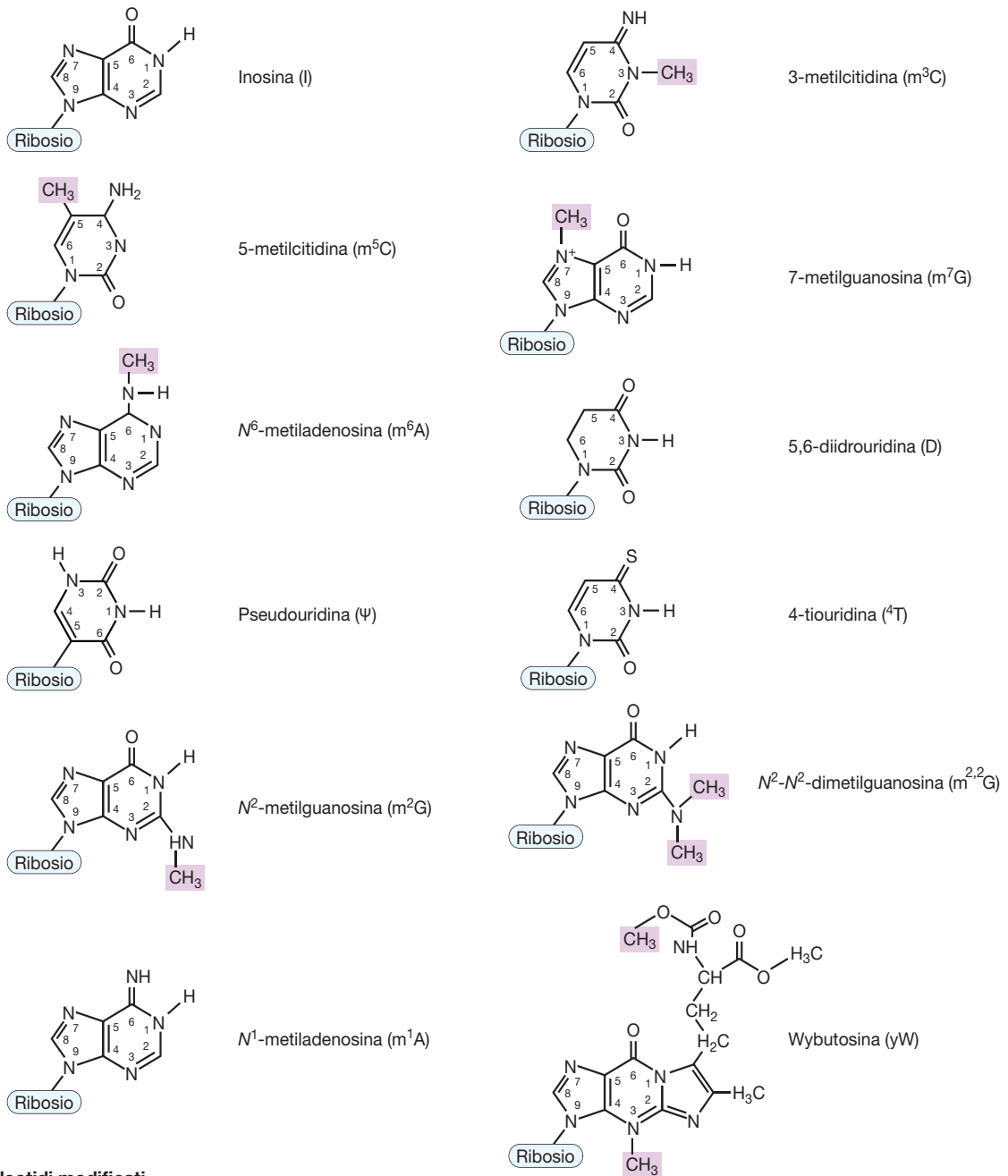
differiscono per un gruppo OH presente sulla posizione 2 del ribosio e che manca nel deossiribosio. Come vedremo più avanti, questa differenza è molto rilevante, in quanto questo gruppo OH conferisce instabilità alle molecole di RNA. Il DNA che ha un H in quella stessa posizione è metabolicamente più stabile.

Le basi azotate che si trovano negli acidi nucleici naturali sono di due tipi: **purine** (a doppio anello eterociclico) e **pirimidine** (a singolo anello). La figura 2.1 mostra le specifiche basi che si trovano negli acidi nucleici naturali: due purine, adenina e guanina, presenti sia nel DNA che nell'RNA, e tre tipi di pirimidine, citosina, timina e uracile, di cui la citosina è comune a DNA e RNA, mentre la timina è componente del solo DNA trovandosi al suo posto l'uracile nell'RNA. La differenza tra uracile e timina è la presenza di un gruppo metilico in posizione 5 dell'anello pirimidinico. Di solito ci si riferisce alle basi con le loro iniziali A, C, G, T e U.

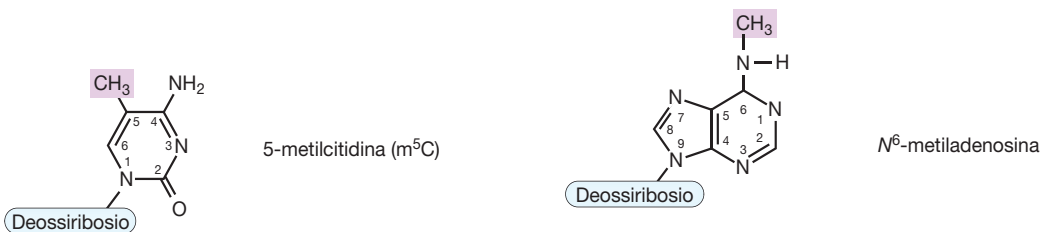
La **figura 2.2** illustra le posizioni e le modificazioni chimiche che possono avvenire nelle basi. Più avanti vedremo come queste basi modificate sia nell'RNA che nel DNA abbiano un ruolo essenziale in molti processi biologici.

Quando una delle suddette basi è legata alla posizione 1 di uno zucchero (ribosio o deossiribosio), abbiamo i **nucleosidi**, che prendono il nome di adenosina, guanosina, citidina, timidina e uridina (**tab. 2.1**). Poiché in ciascun nucleoside sia lo zucchero che la base azotata hanno una numerazione 1, 2, 3 ecc. per indicare le varie posizioni della molecola (*vedi* numerazioni in **fig. 2.3**), si usa aggiungere un apice ['] alla numerazione relativa allo zucchero per distinguerla da quella della base.

Ai nucleosidi possono essere legati uno o più gruppi fosfato e, in tal caso, prendono il nome di **nucleotidi**: questi sono chiamati acido adenilico, acido guanilico ecc., come mostrato in tabella 2.1. Infatti, il gruppo fosfato conferisce una valenza acida alla molecola. La **figura 2.3** mostra come il gruppo fosfato possa trovarsi legato al carbonio 5' dello zucchero oppure al carbonio 3'. Inoltre, considerati come molecole indipendenti (cioè non inserite nella catena polinucleotidica), i nucleotidi possono contenere un solo gruppo fosfato oppure due o tre. In particolare, i nucleotidi che vengono utilizzati come substrato per la sintesi del DNA e dell'RNA hanno tre gruppi fosfato legati in serie sulla posizione 5' dello zucchero, come mostrato in **figura 2.4**. In questo caso i tre gruppi fosfato vengono indicati, a partire da quello legato direttamente allo zucchero, con α , β e γ . Come mostrato in tabella 2.1, l'abbreviazione per i nucleotidi è la lettera che indica la base (A, G ecc.) preceduta da una lettera "d" minuscola nel caso la base sia legata al deossiribosio, e seguita dall'indicazione del numero di gruppi fosfato: MP (monofosfato), DP (difosfato) o TP (trifosfato). Per esempio: dGMP, deossiguanosina monofosfato, è un deossiribosio legato a una guanina e a un gruppo fosfato; ATP, adenosina trifosfato, è un ribosio legato a un'adenina e a tre gruppi fosfato. In **figura 2.5** si può vedere che una base legata allo zucchero si definisce nucleoside e se il nucleoside è legato al fosfato si chiama nucleotide.



Ribonucleotidi modificati



Deossiribonucleotidi modificati

Figura 2.2 Basi modificate che si possono trovare nell'RNA e nel DNA.

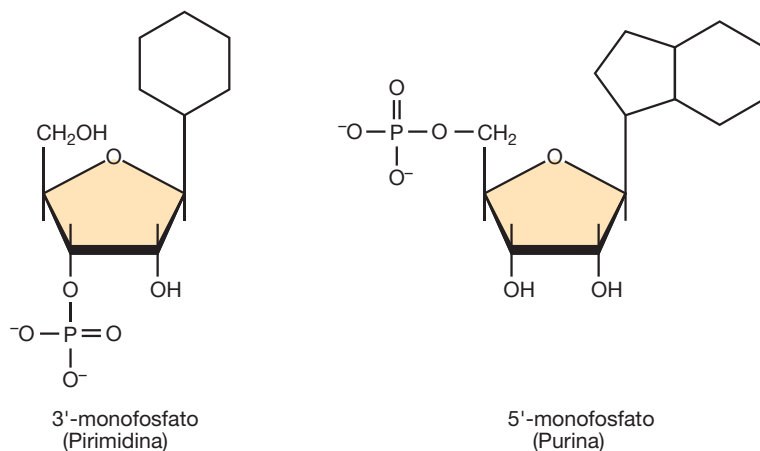


Figura 2.3 I nucleotidi possono portare il gruppo fosfato in posizione 5' o 3'.

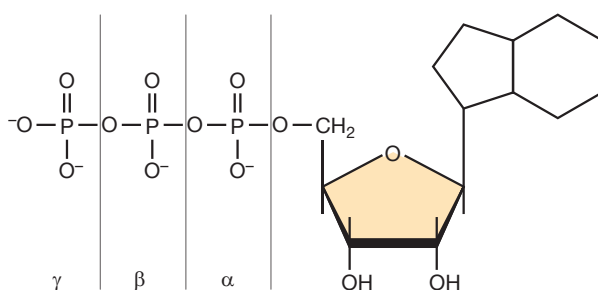


Figura 2.4 Nomenclatura dei gruppi fosfato in un nucleoside trifosfato. Il gruppo fosfato in posizione α è quello che si ritrova nello scheletro degli acidi nucleici.

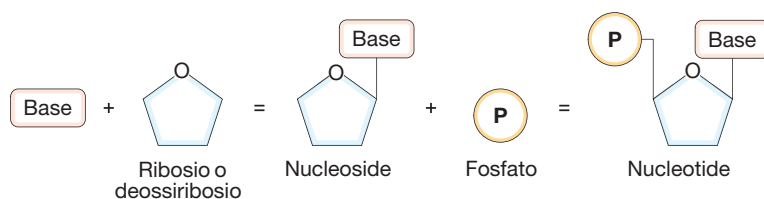
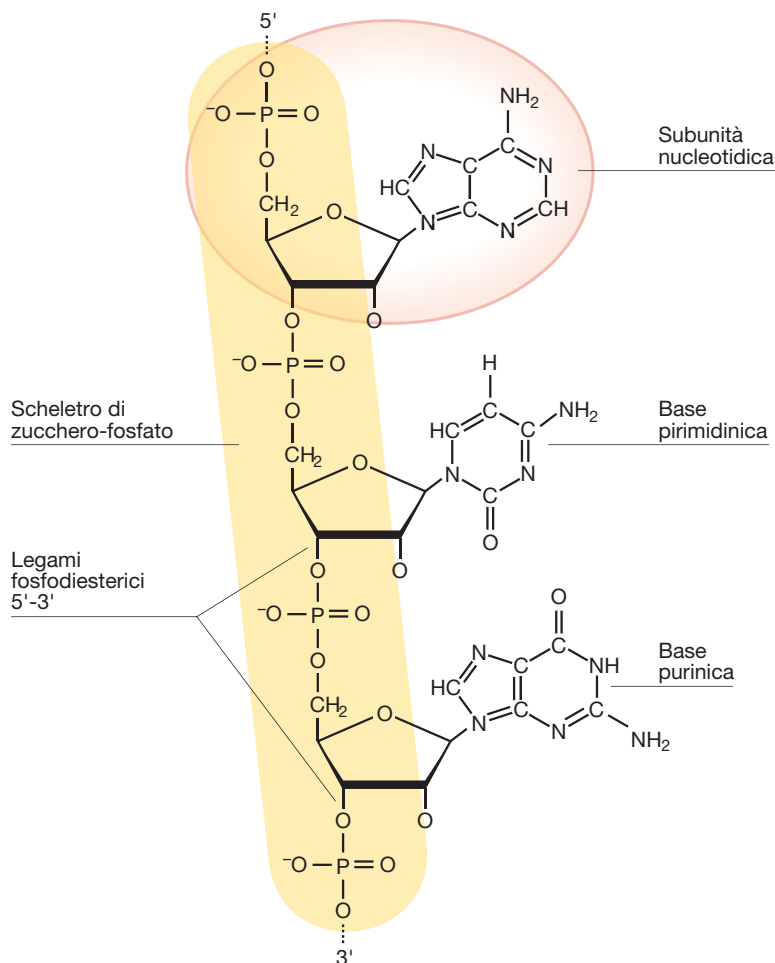


Figura 2.5 Basi, nucleosidi e nucleotidi. Una base legata allo zucchero costituisce un nucleoside. Se il nucleoside è legato a uno o più gruppi fosfato viene detto nucleotide.

Passiamo ora a vedere come i nucleotidi sono legati tra loro a formare una catena polinucleotidica di DNA o di RNA. La **figura 2.6** mostra che lo scheletro (o impalcatura) della catena polinucleotidica è costituito dall'alternanza di zuccheri (deossiribosio o ribosio) e di gruppi fosfato, e che le basi sporgono lateralmente da questo scheletro. Infatti, ciascuna base è legata alla posizione 1' di uno zucchero da un legame glicosidico che interessa l'N1 delle pirimidine o l'N9 delle purine. Va notato che tra uno zucchero e l'altro c'è un solo gruppo fosfato e che quindi si tratta di un polimero di nucleosidi monofosfato; infatti, durante la sintesi degli acidi nucleici, ciascun nucleoside trifosfato utilizzato come substrato, perde due dei suoi tre gruppi fosfato (*vedi oltre*). Come evidenziato nella figura, ciascun gruppo fosfato forma un legame estere con il carbonio 5' di uno zucchero e un secondo legame estere con il carbonio 3' dello zucchero successivo. Questo tipo di **legame**, chiamato **fosfodiesterico**, conferisce una sorta di asimmetria, o meglio di polarità, alla catena polinucleotidica, nel senso che questa presenta due diverse estremità: da una parte c'è un nucleotide che ha il carbonio 5' libero, mentre il carbonio 3' è impegnato nel legame fosfodiesterico con il nucleotide adiacente; dall'altra la molecola polimerica termina con un nucleotide che ha il carbonio 5' impegnato nel legame fosfodiesterico con quello adiacente, mentre

Figura 2.6 Una catena polinucleotidica ha una struttura ripetitiva. Essa consiste di uno scheletro (o impalcatura) in cui si alternano zuccheri e gruppi fosfato, legati tra loro da legami 5' → 3' fosfodiesterici. Da questa impalcatura protrudono lateralmente le basi azotate.



il carbonio 3' è libero. 5' e 3' liberi vuol dire che questi C-terminali non sono impegnati in legami fosfodiesterici; la loro valenza è comunque impegnata da gruppi -OH o da gruppi fosfato.

Questa polarità 5' → 3' delle catene polinucleotidiche, sia di DNA che di RNA, è molto importante. Infatti, la struttura dello scheletro della catena polinucleotidica è diversa se vista da un estremo o dall'altro e, come vedremo nelle pagine seguenti, tutte le reazioni di sintesi e copiatura (replicazione, trascrizione e traduzione), e anche la maturazione e la degradazione degli acidi nucleici, coinvolgono attività enzimatiche che sono specifiche per l'una o l'altra direzione.

Per convenzione le sequenze degli acidi nucleici si scrivono in direzione 5' → 3', cioè dall'estremità 5' libera a sinistra all'estremità 3' libera a destra. In disegni, schemi e sequenze è in genere utile indicare le due estremità, ma è sicuramente indispensabile evidenziarle laddove non fosse rispettata la direzione convenzionale 5' → 3'.

La **figura 2.7** mostra un modo spesso usato per schematizzare una catena polinucleotidica: in questo schema le barrette verticali indicano gli zuccheri e quelle oblique i gruppi fosfato che uniscono il carbonio 3' di uno zucchero con il carbonio 5' dello zucchero adiacente. Si noti che a sinistra vi è l'estremità 5' libera (in questo caso porta un gruppo fosfato) e a destra l'estremità 3' libera (in questo caso porta un gruppo OH). Le basi vengono indicate con le convenzionali lettere A, U, C, G, legate al carbonio 1' degli zuccheri.

Negli anni quaranta del secolo scorso fu acquisita un'altra informazione sulla struttura chimica del DNA. Si tratta delle ricerche del chimico di origine austriaca

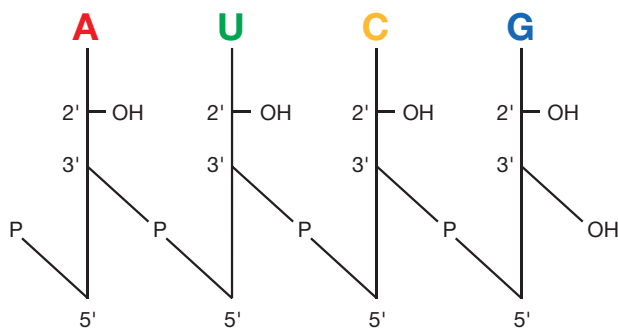


Figura 2.7 Schema semplificato di una catena polinucleotidica di RNA.

Tabella 2.2 Composizione in basi del DNA di varie specie.

Fonte	A	G	C	T	A/T	G/C	G+C	Py/Pu
<i>Escherichia coli</i>	26,0	24,9	25,2	23,9	1,08	0,99	50,1	1,04
<i>Mycobacterium tuberculosis</i>	15,1	34,9	35,4	14,6	1,03	0,99	70,3	1,00
Lievito	31,7	18,3	17,4	32,6	0,97	1,05	35,7	1,00
Bue	29,0	21,2	21,2	28,7	1,01	1,00	42,4	1,01
Maiale	29,8	20,7	20,7	29,1	0,92	1,00	41,4	1,01
Uomo	30,4	19,9	19,9	30,1	1,01	1,00	39,8	1,01

Erwin Chargaff (1905-2002), anche lui emigrato per motivi razziali negli Stati Uniti, dove condusse, presso la Columbia University a New York, le sue ricerche sulla composizione in basi del DNA. Chargaff si era accorto che preparazioni di DNA di origine diversa possono avere un differente contenuto delle quattro basi A, C, G e T (esprese come contenuto percentuale). Compì quindi numerosi e lunghi studi di analisi quantitativa e concluse che, mentre organismi di diverse specie possono avere DNA con composizione in basi differente, il DNA preparato da diversi tessuti od organi della stessa specie ha sempre la stessa composizione in basi. Questa costanza nell'ambito dello stesso organismo e diversità tra organismi diversi ben si accordava con l'idea del ruolo del DNA come materiale genetico. Nel corso di questi studi Chargaff fece un'altra osservazione che si rivelò poi importante nel momento in cui Crick e Watson affrontarono il problema della struttura fisica del DNA. Come mostrato in **tabella 2.2**, Chargaff osservò che, qualsiasi sia la fonte del DNA e nonostante la diversa composizione in basi, veniva sempre rispettata una regola (che venne poi chiamata **regola di Chargaff**): la percentuale di A è sempre uguale alla percentuale di T, e la percentuale di C è sempre uguale alla percentuale di G. Questa regola non è valida nel caso dell'RNA.

2.2 Struttura fisica del DNA: la scoperta della struttura a doppia elica

Era chiaro in quegli anni che per svelare il mistero della funzione dei geni fosse necessario conoscere la struttura del DNA. Per comprendere la funzione di una macromolecola, infatti, non basta conoscerne la struttura chimica, ma è necessario conoscerne la struttura tridimensionale. È interessante il fatto che già nel 1940 due personaggi di primo piano della scienza di allora, Max Delbrück (*vedi* fig.1.2) e Linus Pauling (**fig. 2.8**), scrissero insieme un lavoro pubblicato sulla rivista scientifica *Science* in cui si avventuravano in congetture sulla struttura del gene. A quell'epoca non si sapeva ancora che il materiale genetico fosse il DNA. I

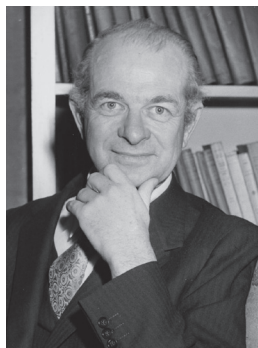


Figura 2.8 Linus Pauling (1901-1994). Pauling è stato un grande chimico che oltre a risolvere la struttura dell'alfa elica delle proteine, ricerca per la quale ha ottenuto il premio Nobel, ha svolto studi fondamentali sui legami chimici. Ha ricevuto anche un secondo Nobel per la pace.

due autori pensavano piuttosto a geni fatti di proteine, tuttavia raggiunsero delle conclusioni che potremmo definire profetiche.

Il loro ragionamento partiva dalla considerazione che una delle due proprietà fondamentali di un gene è la capacità di essere duplicato (l'altra è la proprietà di esprimere un carattere o fenotipo). La duplicazione di un oggetto non è qualche cosa che in natura si osserva comunemente, se si escludono i sistemi viventi, dove appunto sono implicati i geni. Anche con le sofisticate tecnologie di oggi, non si è riusciti a costruire una macchina che sappia duplicarsi. Al massimo possiamo fare una macchina molto complessa capace di duplicare un oggetto semplice. Quello che invece appare molto più facile è fare uno stampo (un calco), producendo quindi un oggetto con struttura complementare all'oggetto originale. Con questo stampo possiamo, ripetendo l'operazione, riprodurre l'oggetto originale. Applicando questo ragionamento alla struttura dei geni, Delbrück e Pauling conclusero che, strutturalmente, un gene doveva essere costituito da due parti complementari, nel senso di superfici complementari che possano fare ciascuna da stampo per la replicazione. È evidente come questa idea si sia poi rivelata esatta nella struttura a due filamenti complementari del DNA.

Dopo l'identificazione del DNA come materiale genetico, vari laboratori cominciarono a interessarsi alla sua struttura tridimensionale utilizzando per lo più le tecniche che erano state messe a punto e già applicate con successo allo studio della struttura delle proteine. Tra questi pochi gruppi primeggiavano, agli inizi degli anni cinquanta, due laboratori: quello di Linus Pauling, in California, e quello di Maurice Wilkins e Rosalind Franklin a Londra (**fig. 2.9A**). La tecnica che si è dimostrata fondamentale per risolvere la struttura del DNA è stata la **diffrazione dei raggi X**, anche se altri tipi di informazioni ottenute con metodi diversi sono state pure importanti.

La tecnica della diffrazione dei raggi X richiede che la macromolecola che si vuole analizzare possa essere ottenuta in forma di cristallo in cui le molecole, e quindi gli atomi che le compongono, sono tutte allineate in maniera ordinata. La **figura 2.10** mostra uno schema semplificato di un diffrattometro a raggi X. Mentre era relativamente facile ottenere piccoli cristalli di proteine purificate, questo non si rivelò possibile per il DNA. Tuttavia, anche se non si tratta di un vero cristallo, il DNA può essere ridotto in forma di fibre in cui le lunghe molecole del DNA sono orientate più o meno parallelamente l'una all'altra. Queste fibre di molecole orientate permisero di ottenere delle immagini di diffrazione dei raggi

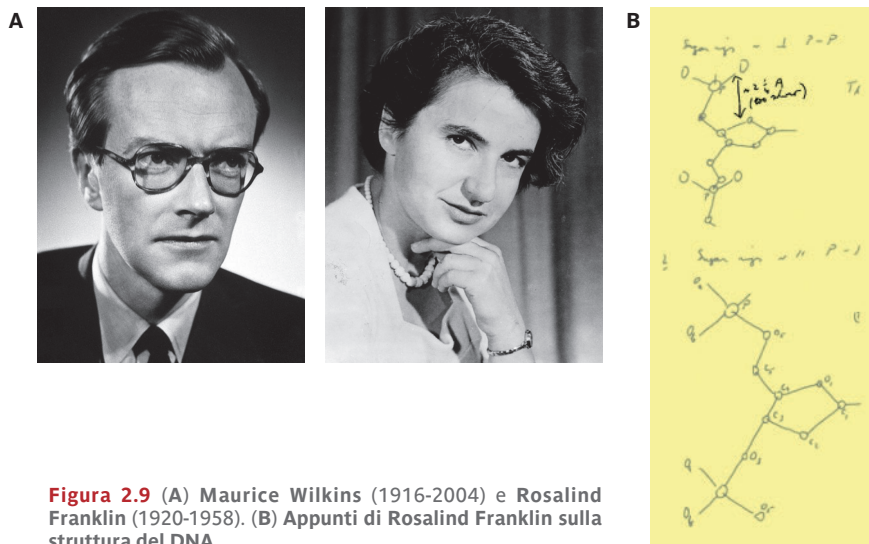


Figura 2.9 (A) Maurice Wilkins (1916-2004) e Rosalind Franklin (1920-1958). (B) Appunti di Rosalind Franklin sulla struttura del DNA.

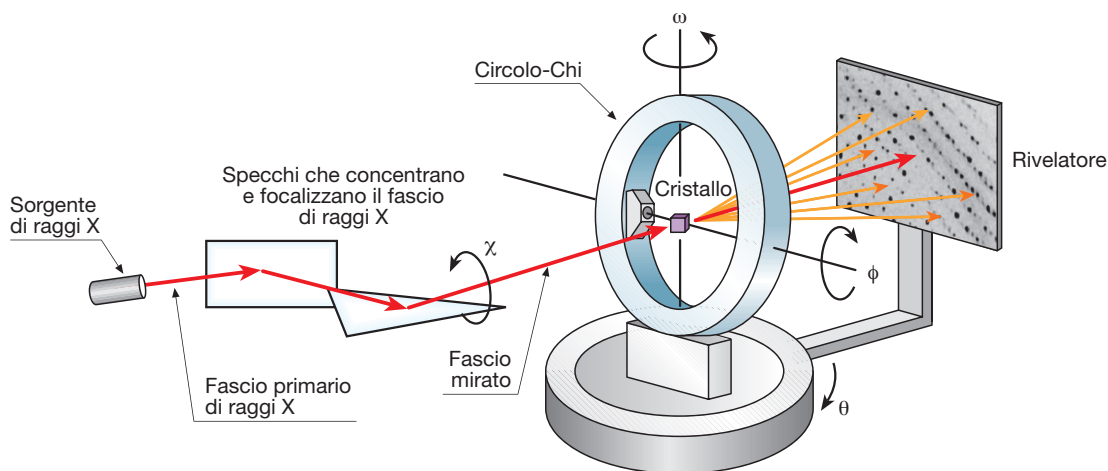


Figura 2.10 Schema semplificato di un diffrattometro a raggi X.

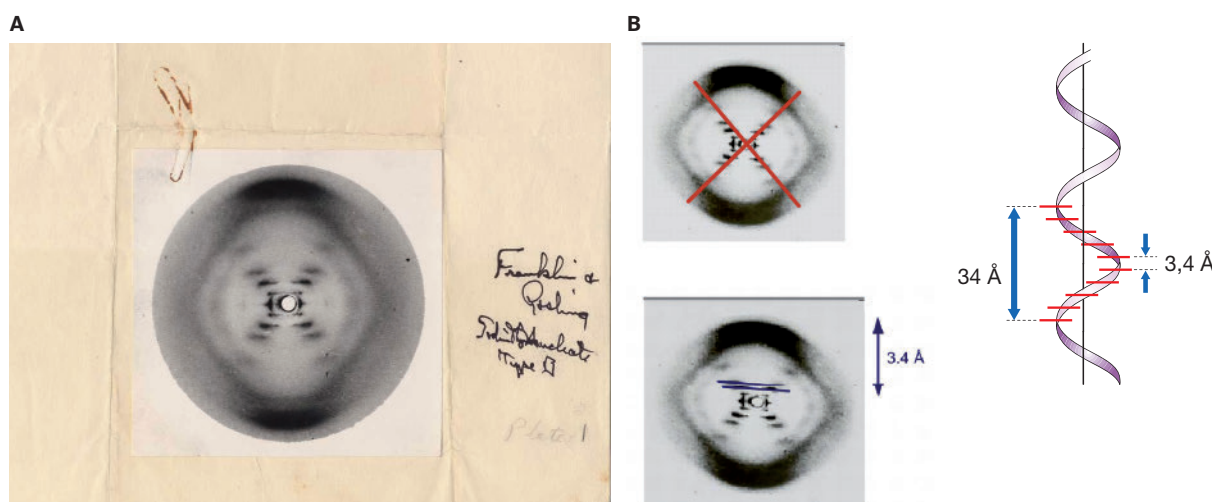
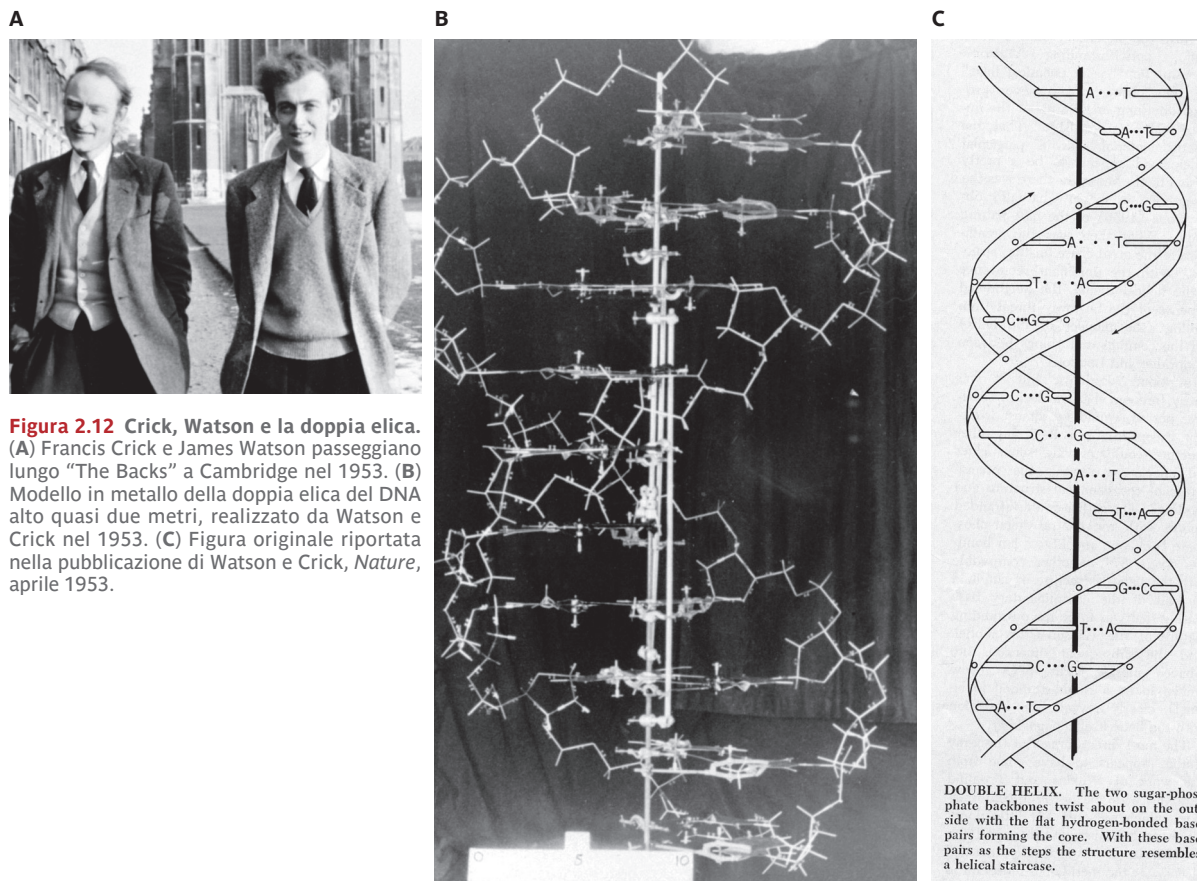


Figura 2.11 Fotografie di diffrazione dei raggi X di fibre di DNA. (A) Fotografia originale ottenuta da Rosalind Franklin. (B) Appunti per il calcolo dei parametri dell'elica effettuati da Francis

Crick sulla base della foto di diffrazione prodotta dalla Franklin. (Per gentile concessione del Dolan DNA Learning Center, Cold Spring Harbor, NY.)

X che, anche se non ottimali, consentirono di ricavare alcuni parametri strutturali. Nel 1952 Pauling propose una struttura, che però si rivelò poi sbagliata: tre eliche intrecciate tra loro con lo scheletro zucchero-fosfato all'interno e le basi sporgenti verso l'esterno. Questo particolare delle basi rivolte all'esterno era stato proposto per giustificare il fatto che i dati sperimentali indicavano che gli scheletri zucchero-fosfato delle eliche formano una struttura regolare con diametro abbastanza costante, cosa difficile da spiegare se le basi, pirimidine più piccole e purine più grandi, si fossero trovate all'interno della struttura. Nel frattempo Rosalind Franklin a Londra andava producendo immagini di diffrazione dei raggi X sempre migliori (**fig. 2.11A**) ma evitava di lanciarsi in proposte di modelli ritenendo necessario accumulare più dati certi. È a questo punto che intervennero Francis Crick e James Watson (**fig. 2.12A**). Crick era già al Cavendish Laboratory a Cambridge e lavorava a problemi di struttura delle proteine. Watson era un



giovannissimo borsista venuto dagli Stati Uniti, dove aveva fatto il dottorato con Salvador Luria, e proveniva, quindi, dal menzionato gruppo del fago. I due stabilirono una collaborazione, oltre a un'amicizia personale, e avendo ben chiaro che scoprire la struttura del DNA era la vera sfida del momento che avrebbe cambiato completamente la Biologia, si dedicarono a questo problema. I dati sperimentali che utilizzavano erano, oltre a quelli prodotti da loro su una fibra di DNA diversa da quella analizzata dalla Franklin, quelli che gli altri gruppi, soprattutto i vicini del King's College a Londra, pubblicavano o presentavano in occasione di convegni o seminari scientifici. Quando Crick vide la foto di diffrazione prodotta dalla Franklin, capì subito che si trattava di un'elica e provò a calcolarne i parametri, come mostrato in **figura 2.11B**. Inoltre, Watson e Crick scelsero come approccio quello di tentare di costruire dei modelli (con le basi ritagliate in cartone, fili di ferro ecc.) e di confrontarli con i dati sperimentali di diffrazione dei raggi X, procedendo in una serie di prova-errore. Watson racconta che la mattina del 28 febbraio 1953 stava provando con dei modelli di cartone a formare possibili appaiamenti tra le basi e si accorse che se si appaiava A con T e G con C si ottenevano due coppie le cui estremità avevano la stessa distanza. La cosa più importante era che le coppie così formate erano sovrapponibili e, quindi, in grado di formare molecole di lunghezza praticamente infinita se si impilavano una sull'altra. Inoltre, erano convinti che la struttura di Pauling con i fosfati al centro dell'elica fosse sbagliata e quindi decisero di disporre i due filamenti zucchero-fosfato all'esterno e di sistemare le basi affacciate all'interno della struttura. Questa soluzione, insieme all'idea di mettere di fronte una A a una T e una C a una G, permetteva allo stesso tempo di giustificare la regolarità del diametro della **doppia elica** (ogni coppia di

basi è costituita da una purina e una pirimidina) e di dare una spiegazione logica alla regola di Chargaff (la quantità di A è uguale alla quantità di T e la quantità di G è uguale alla quantità di C). La struttura del DNA era dunque risolta. La **figura 2.12B** mostra il modello di metallo che avevano costruito per studiarne le caratteristiche, mentre la **figura 2.12C** mostra la figura originale dell'articolo che Watson e Crick pubblicarono sulla rivista scientifica *Nature* il 25 aprile 1953.

2.3 Struttura fisica del DNA: i parametri strutturali della doppia elica

Secondo il modello proposto da Crick e Watson le coppie di basi, che con i loro anelli eterociclici sono strutture sostanzialmente piatte, sono disposte quasi perpendicolarmente rispetto all'asse della doppia elica, come a formare dei gradini di una scala a chiocciola (**fig. 2.13**). Le basi che si affacciano in ciascuna coppia sono sempre una pirimidina e una purina e in particolare vi è sempre l'appaiamento tra una A e una T e tra una G e una C, il che giustifica la regolarità del diametro della doppia elica. Come mostrato in **figura 2.14**, nella coppia di basi A-T la A si viene a trovare di fronte alla T con una geometria e una distanza tali da permettere la formazione di due legami idrogeno, mentre nella coppia G-C la G e la C formano tre legami idrogeno. Una caratteristica importante del modello è che le due catene polinucleotidiche sono "antiparallele", cioè sono disposte con una polarità 5' → 3' opposta l'una rispetto all'altra, come mostrato in **figura 2.15** dove si può osservare anche come i due filamenti siano costituiti dall'alternanza di residui di deossiribosio e di gruppi fosfato, mentre le basi sporgono lateralmente per appaiarsi con quelle dell'altro filamento.

La doppia elica (duplex) del DNA ha una struttura regolare, destrorsa, compie un giro completo ogni 34 Å e ha un diametro di circa 20 Å. La distanza tra due coppie

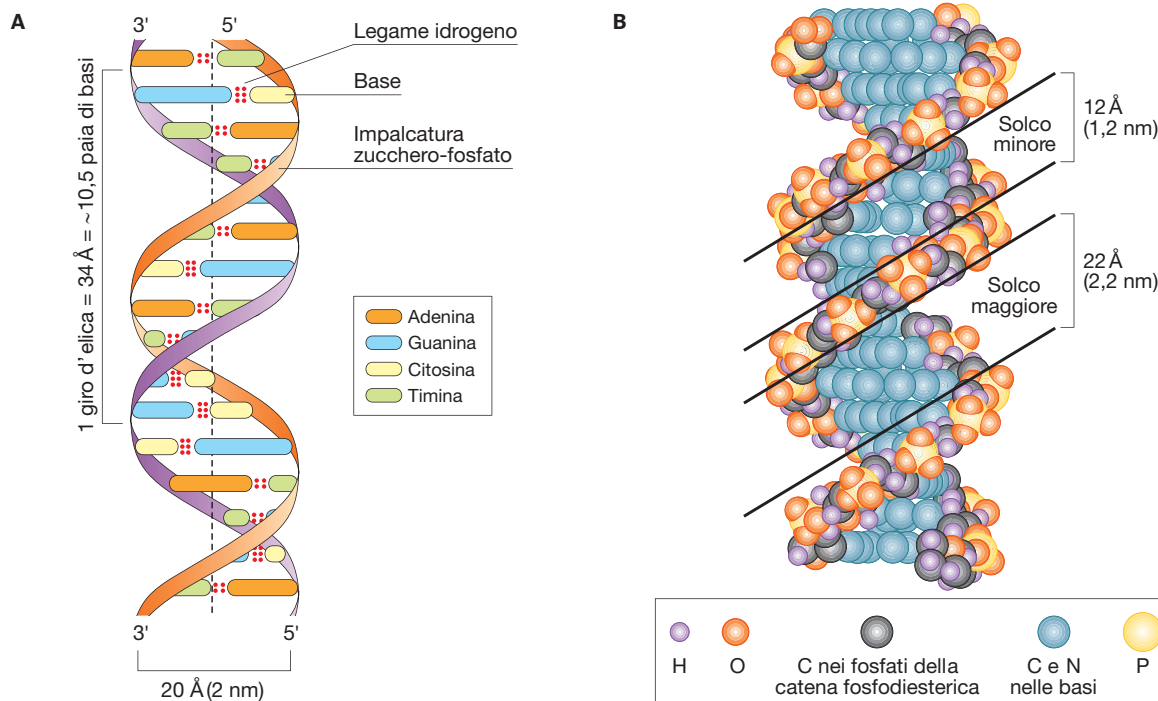


Figura 2.13 Due diverse rappresentazioni grafiche della doppia elica del DNA, dove sono indicate le dimensioni e alcune caratteristiche strutturali. (A) Rappresentazione schematica in

cui l'impalcatura zucchero-fosfato è rappresentata dai nastri. (B) Struttura molecolare ad atomi pieni.

Figura 2.14 Legami idrogeno tra le coppie di basi del DNA. (A) Tra la G e la C della coppia di basi G-C si formano tre legami idrogeno. (B) Tra la A e la T della coppia A-T se ne formano due.

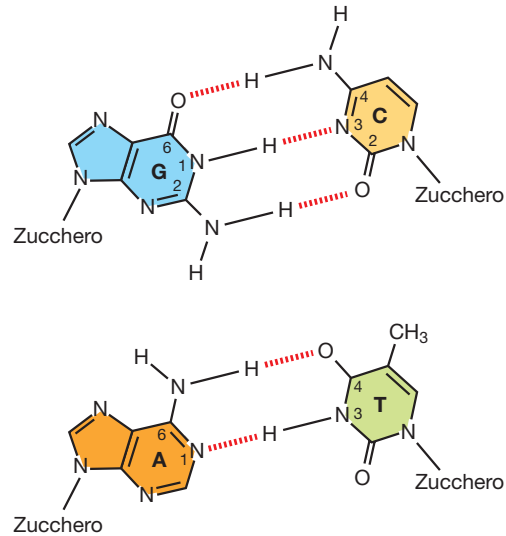
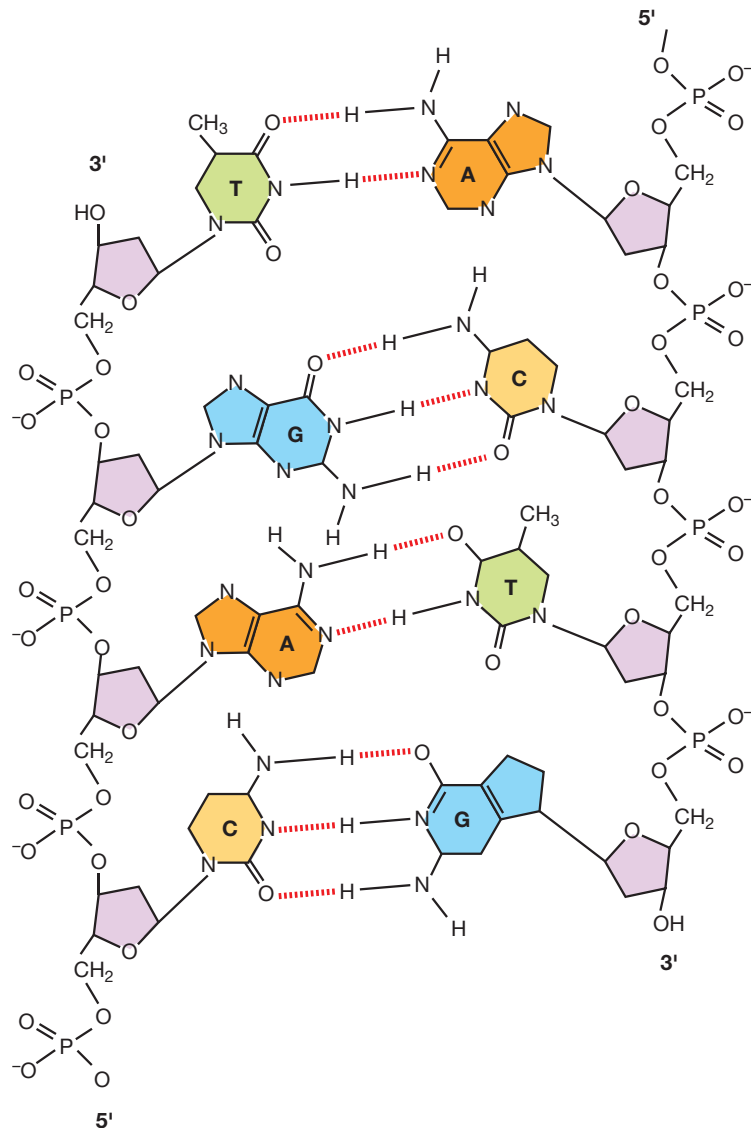


Figura 2.15 Appaiamento tra le basi di due catene polinucleotidiche. Si noti l'appaiamento specifico tra le basi complementari (A con T e G con C) e l'opposta polarità 5' → 3' delle due catene polinucleotidiche complementari.



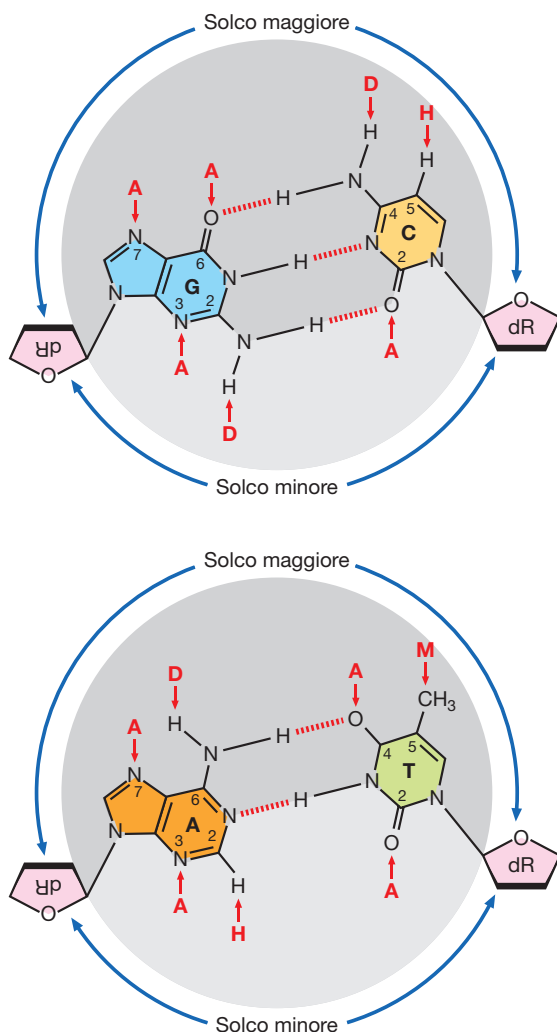


Figura 2.16 Gruppi chimici delle basi esposti nel solco maggiore e nel solco minore del DNA. Si noti la posizione di legame delle basi con gli zuccheri, spostata rispetto all'asse centrale delle basi appaiate. Questa posizione crea un'asimmetria nella molecola generando i due solchi di dimensioni diverse. Le lettere in rosso identificano i vari gruppi: A: accettore di legami idrogeno; D: donatore di legame idrogeno; H: idrogeni non polari; M: gruppo metilico.

di basi adiacenti è di 3,4 Å e ci sono, quindi, circa 10 coppie di basi per ogni giro di elica. L'elica presenta un'asimmetria dovuta alla posizione delle molecole di deossiribosio ai lati delle basi: infatti, come conseguenza dell'opposta polarità 5' → 3' delle due eliche, i due zuccheri di ciascuna coppia di nucleotidi si vengono a trovare dallo stesso lato (**fig. 2.16**). Come evidente in figura 2.13, questa asimmetria genera nella doppia elica due solchi di dimensioni diverse, detti appunto **solco maggiore** e **solco minore**. Questi due solchi sono molto importanti per il riconoscimento della sequenza del DNA: il solco maggiore ha molti più atomi che possono essere donatori o accettori di legami idrogeno e come conseguenza si può dire che il “linguaggio” presentato da questa zona dell'elica è molto più vario e ricco che nel solco minore. Inoltre, come mostrato in **figura 2.17**, l' α -elica delle proteine ha la dimensione giusta per poter interagire esattamente con il solco maggiore. Come vedremo in seguito, infatti, quasi tutte le proteine che si legano specificamente al DNA hanno dei domini ad α -elica che riconoscono le sequenze esposte nel solco maggiore.

Stabilità della doppia elica di DNA in soluzione

Facciamo qualche considerazione sulla stabilità della molecola di DNA a doppia elica in soluzione acquosa. Le catene polinucleotidiche singole hanno una parte idrofila, lo scheletro zucchero-fosfato, e una parte idrofobica, le basi. Nella

Figura 2.17 Le proteine interagiscono generalmente con il solco maggiore del DNA. Il solco maggiore della doppia elica del DNA ha la dimensione giusta per accogliere e interagire con la struttura α -elica delle proteine.

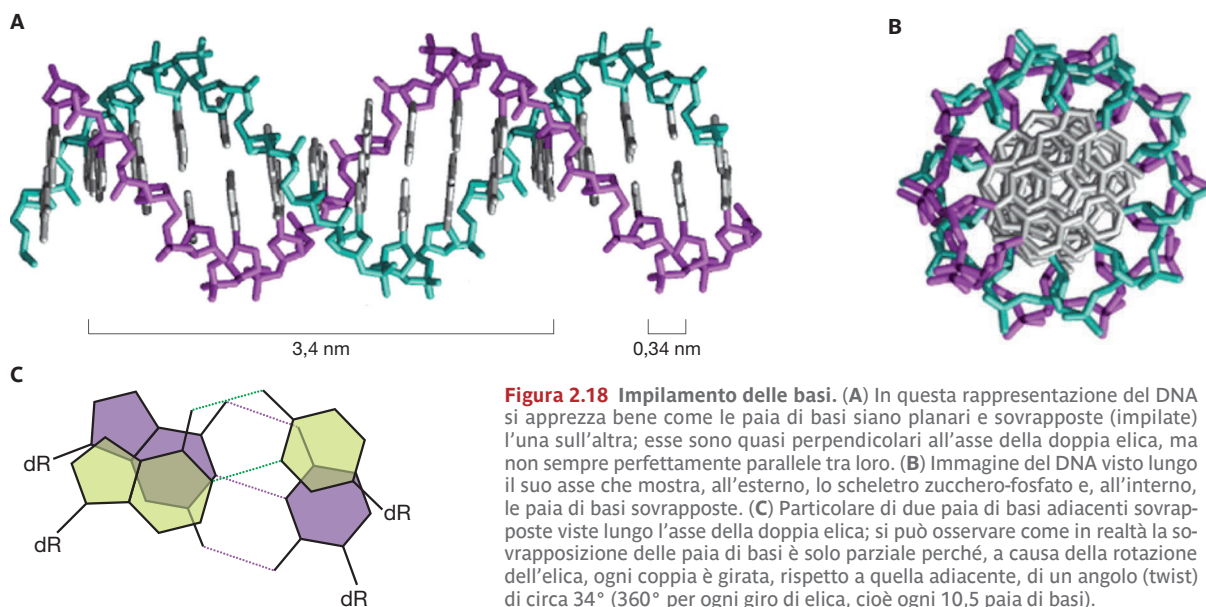
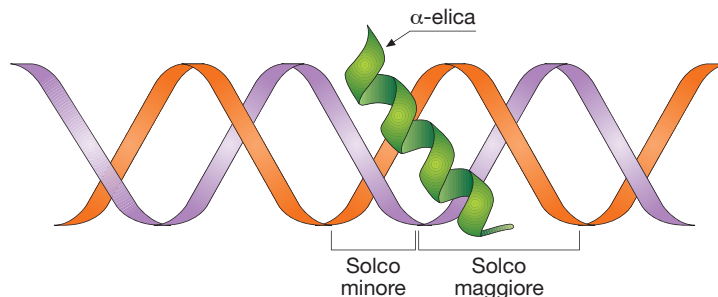


Figura 2.18 Impilamento delle basi. (A) In questa rappresentazione del DNA si apprezza bene come le paia di basi siano planari e sovrapposte (impilate) l'una sull'altra; esse sono quasi perpendicolari all'asse della doppia elica, ma non sempre perfettamente parallele tra loro. (B) Immagine del DNA visto lungo il suo asse che mostra, all'esterno, lo scheletro zucchero-fosfato e, all'interno, le paia di basi sovrapposte. (C) Particolare di due paia di basi adiacenti sovrapposte viste lungo l'asse della doppia elica; si può osservare come in realtà la sovrapposizione delle paia di basi è solo parziale perché, a causa della rotazione dell'elica, ogni coppia è girata, rispetto a quella adiacente, di un angolo (twist) di circa 34° (360° per ogni giro di elica, cioè ogni 10,5 paia di basi).

struttura a doppia elica i due scheletri zucchero-fosfato, idrofilici, con le cariche negative dei fosfati, si dispongono all'esterno della struttura a contatto con l'acqua, mentre le basi (idrofobiche) vanno a disporsi all'interno sfuggendo così all'acqua. Nel suo insieme, la stabilità della doppia elica, cioè la propensione a mantenere le due catene polinucleotidiche legate tra loro, dipende da varie forze. I due scheletri che contengono gli atomi di fosforo carichi negativamente tendono a respingersi; questa repulsione viene controbilanciata da due tipi di forze che tendono a mantenerle appaiate. Innanzitutto i legami idrogeno tra le basi, che, come abbiamo detto, sono tre tra ogni coppia G-C e due tra ogni coppia A-T. Benché questo appaiamento delle basi complementari sia essenziale per la specificità di legame, i legami H tra le basi contribuiscono solo in parte alla stabilità della doppia elica. Infatti, questa stabilità è aumentata dall'effetto idrofobico dovuto all'**impilamento** (*stacking*) delle basi. Questo è spiegato in **figura 2.18**, dove, guardando lungo l'asse dell'elica, si vede come le coppie di basi adiacenti presentino una considerevole sovrapposizione, anche se non completa a causa della rotazione dell'elica. Questa sovrapposizione è stabilizzata dalla idrofobicità delle basi stesse, che, così impilandosi, mantengono minimo il contatto con l'acqua. Gli orbitali π degli anelli aromatici tendono a formare legami deboli che stabilizzano l'elica. La **tabella 2.3** riporta il contenuto energetico delle varie combinazioni tra le coppie di basi. Va sottolineato che c'è una differenza di energia se sopra una coppia di AT c'è una coppia di GC o di CG, nel senso che la geometria di impilamento ha un'influenza da un punto di vista energetico.

Tabella 2.3 Energia di impilamento delle coppie di paia di basi.

Coppie di paia di basi impilate	Energia di impilamento (kcal/mole/coppia di pb)
5' – GC – 3' 3' – CG – 5'	–14,59
5' – AC – 3' 3' – TG – 5'	–10,51
5' – TC – 3' 3' – AG – 5'	–9,81
5' – CG – 3' 3' – GC – 5'	–9,69
5' – GG – 3' 3' – CC – 5'	–8,26
5' – AT – 3' 3' – TA – 5'	–6,57
5' – TG – 3' 3' – AC – 5'	–6,57
5' – AG – 3' 3' – TC – 5'	–6,78
5' – AA – 3' 3' – TT – 5'	–5,37
5' – TA – 3' 3' – AT – 5'	–3,82

Da quanto detto deriva che la stabilità della doppia elica in soluzione non è uguale per tutti i DNA. Infatti, la stabilità dipende dalla composizione in basi del DNA, in quanto un DNA ricco in coppie C-G, con tre legami idrogeno, è più stabile di un DNA ricco in coppie A-T, con solo due legami idrogeno. Inoltre, dipende dalla particolare sequenza di nucleotidi nel DNA: infatti, l'effetto idrofobico dipende dall'impilamento e quindi dalla sovrapposizione di coppie di basi adiacenti e quanto due coppie di basi sono sovrapposte dipende dalla particolare sequenza di nucleotidi. Proprio a causa delle interazioni di impilamento tra le basi e ai gradi di flessibilità che hanno i legami chimici dei nucleotidi, il parallelismo tra due coppie di nucleotidi lungo l'asse dell'elica non è quasi mai perfetto, come si può vedere in **figura 2.18A**, e quindi la relativa posizione delle basi è stata catalogata in vari parametri. Per semplificare ne descriveremo qui solo tre (Twist, Roll e Tilt), che sono illustrati in **figura 2.19**. Il **Twist** rappresenta l'angolo tra due coppie di nucleotidi rispetto all'asse della doppia elica, il **Roll** è l'inclinazione, dovuta alle forze di repulsione tra due coppie di nucleotidi, rispetto all'asse della coppia AT o GC e, infine, il **Tilt** è l'angolo che si forma, sempre per repulsione, rispetto all'asse dello scheletro zucchero-fosfato. Queste posizioni relative, come vedremo in seguito, influenzano la forma stessa del DNA, che può diventare “curvo” in base alla ripetizione di particolari sequenze.

Strutture alternative e strutture superiori degli acidi nucleici

■ Strutture alternative

La struttura a doppia elica proposta da Crick e Watson è quella ottenuta mediante diffrazione dei raggi X di fibre di DNA in condizioni di alta umidità e quindi anche in soluzione acquosa, come è la situazione *in vivo*. Questa struttura, che

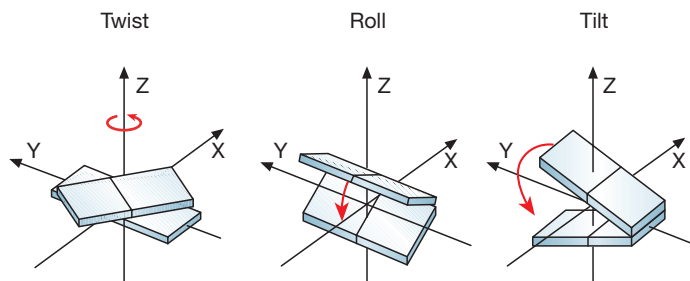
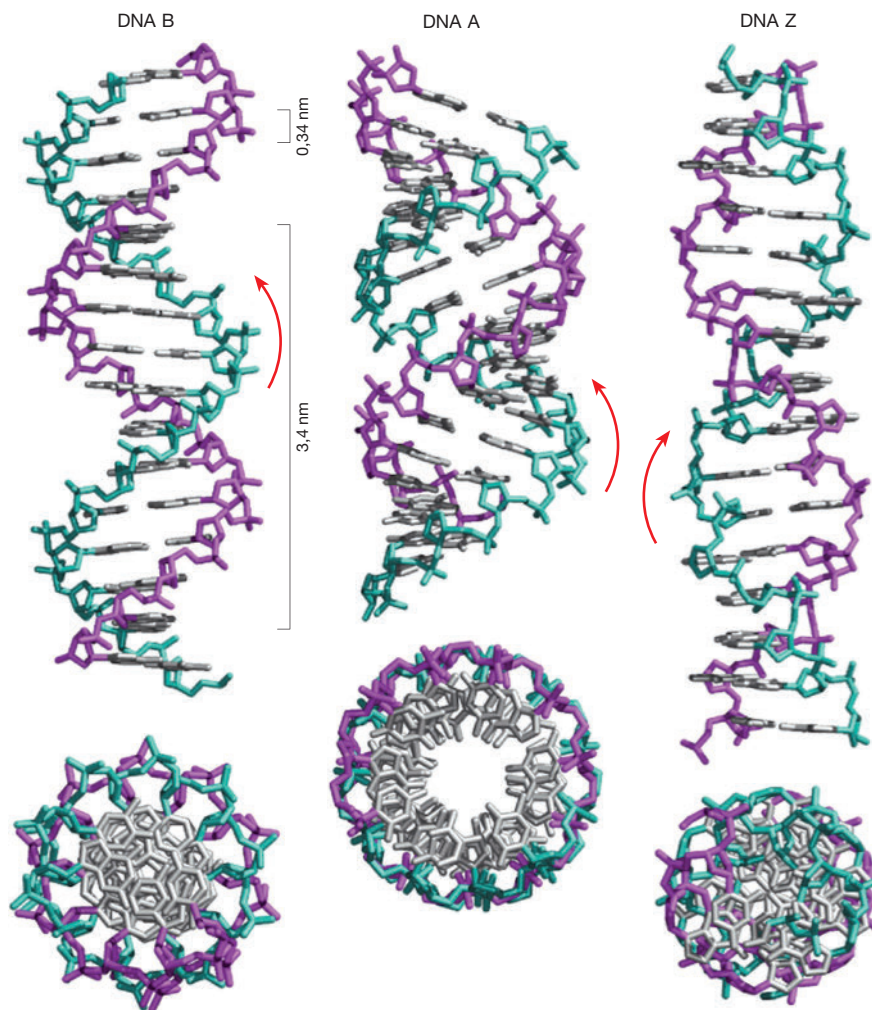


Figura 2.19 I parametri **Twist**, **Roll** e **Tilt** descrivono il relativo posizionamento di due paia di basi adiacenti. Le paia di basi adiacenti impilate nella doppia elica non sono perfettamente parallele, presentando angolazioni che dipendono dalle particolari paia di basi considerate. Queste alterazioni al parallelismo delle basi sono descritte da alcuni parametri. I parametri principali so-

no: il **Twist**, rotazione mediamente di 34° dovuta al giro della doppia elica, e il **Roll** e il **Tilt** che descrivono l'angolatura tra due paia di basi rispetto all'asse dell'elica e allo scheletro zuccherofosfato. Definizione degli assi: Z corrisponde all'asse della doppia elica del DNA; Y indica la direzione dei legami idrogeno tra una coppia di basi; X è l'asse perpendicolare a Y.

Figura 2.20 Strutture alternative del DNA. La doppia elica del DNA può assumere, in particolari condizioni ambientali o per particolari sequenze di basi, varie forme strutturali alternative. Delle varie possibili sono mostrate qui le tre forme di maggior interesse biologico: la forma DNA B è la classica forma del DNA duplex *in vivo*; la forma DNA A si riscontra nel DNA duplex in condizioni artificiali di bassa umidità ma anche, in normali condizioni *in vivo*, nei duplex di RNA e in molecole ibride (eteroduplex) DNA/RNA; la peculiare forma sinistrorsa DNA Z può essere assunta da DNA con sequenza alternata purina/pirimidina, ed è facilitata dal superavvolgimento negativo della molecola di DNA. Le frecce rosse indicano che le doppie eliche in forma A e B sono destrorse, mentre la doppia elica in forma Z è sinistrorsa. (Fonte: Kielkopf C. & Dervan P.B.)



viene identificata come **forma B**, non è l'unica possibile. Infatti, negli anni successivi al lavoro di Crick e Watson, sono state descritte varie altre strutture che la doppia elica di DNA può assumere in varie condizioni ambientali (perlopiù presenza di ioni diversi) o per particolari sequenze di basi. Tralasciamo qui di parlare delle forme C, D, E e altre che sono state descritte per gli acidi nucleici.

Tabella 2.4 Parametri strutturali delle forme A, B e Z del DNA.

Forma	Coppie di basi per giro di elica	Rotazione delle basi (gradi)	Diametro (Å)	Passo dell'elica (Å)	"Rise" (Å per coppia di basi)
A	11	33	23	25	2,3
B	10,5	36	20	34	3,4
Z	12	-30	18	46	3,8

ci, limitandoci a considerare la forma A e la forma Z. La **figura 2.20** mette a confronto le tre strutture del DNA (B, A e Z), mentre la **tabella 2.4** riporta i principali parametri strutturali per queste tre forme alternative.

Se si riduce l'umidità relativa in cui si trova la fibra di DNA, quest'ultimo assume la **forma A**. Questa è destrorsa come la forma B, ma se ne differenzia per vari aspetti:

- le coppie di basi presentano una maggiore angolarità rispetto al piano perpendicolare all'asse della doppia elica;
- ci sono 11 coppie di basi per ogni giro di elica;
- il passo dell'elica è di 25 Å e il diametro è di 23 Å, quindi presenta una forma un po' "tarchiata" rispetto alla forma B.

Mentre la forma B può essere assunta da duplex di DNA solo in condizioni non naturali di bassa umidità, la forma A la si trova anche *in vivo*, quindi in soluzione acquosa, per duplex formati da due filamenti di RNA oppure da un filamento di DNA e uno di RNA. Infatti la presenza dell' -OH nella posizione 2' del ribosio impedisce alla molecola di assumere la forma B.

Di tutte le strutture studiate, la **forma Z** è sicuramente la più particolare. Come si può osservare in figura 2.20, la caratteristica peculiare della forma Z è il fatto di essere un'elica sinistrorsa, anziché destrorsa come tutte le altre forme. I parametri dell'elica comprendono un diametro di 18 Å e un passo di 46 Å, quindi un'elica "magra e allungata" rispetto alla forma B. Questa struttura è stata scoperta studiando DNA sintetici in cui G e C si alternano lungo la sequenza; tuttavia, sono state ottenute prove sperimentali che suggeriscono che una piccola percentuale del DNA presente nelle cellule si trova in forma Z e sono state isolate delle proteine che si legano specificamente a tratti di DNA Z. La causa che genera il DNA Z è il cambiamento di orientamento del legame glicosidico tra la guanina e il deossiribosio. Nella forma B lo zucchero e la base sono presenti nella conformazione "anti", mentre nella forma Z si presentano nella rara conformazione "syn" (**fig. 2.21**). La forma a "zig-zag", da cui il nome Z, si spiega con l'alternanza di conformazioni syn e anti di nucleotidi contigui. La **figura**

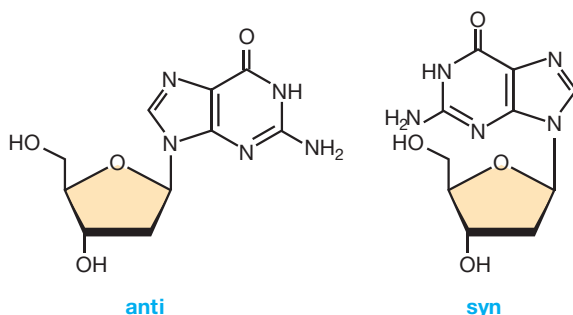


Figura 2.21 Le due posizioni syn e anti della deossiguanosina nelle forme A e Z del DNA. Nel DNA a elica destrorsa il legame che unisce il deossiribosio alla base è sempre in forma anti. Nel DNA Z il legame ruota facendo assumere alla guanosina una conformazione syn. Nelle pirimidine, il legame è generalmente nella conformazione anti per l'interferenza sterica tra lo zucchero e l'ossigeno carbonilico nella posizione 2 delle pirimidine. Quindi, in una coppia G-C solo la G potrà rovesciare la posizione dello zucchero.

Figura 2.22 Nella stessa molecola di DNA possono coesistere tratti di DNA B e di DNA Z. (A) Un tratto di DNA Z è intercalato in una molecola a struttura DNA B. (B) Dettaglio del punto di transizione dalla forma Z alla forma B.

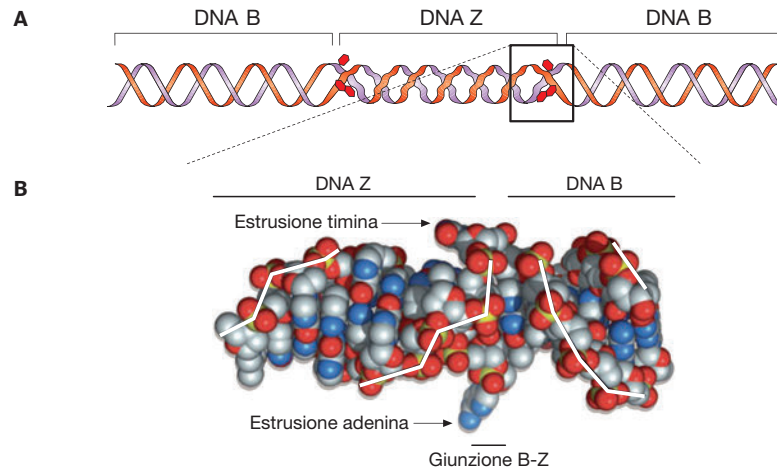
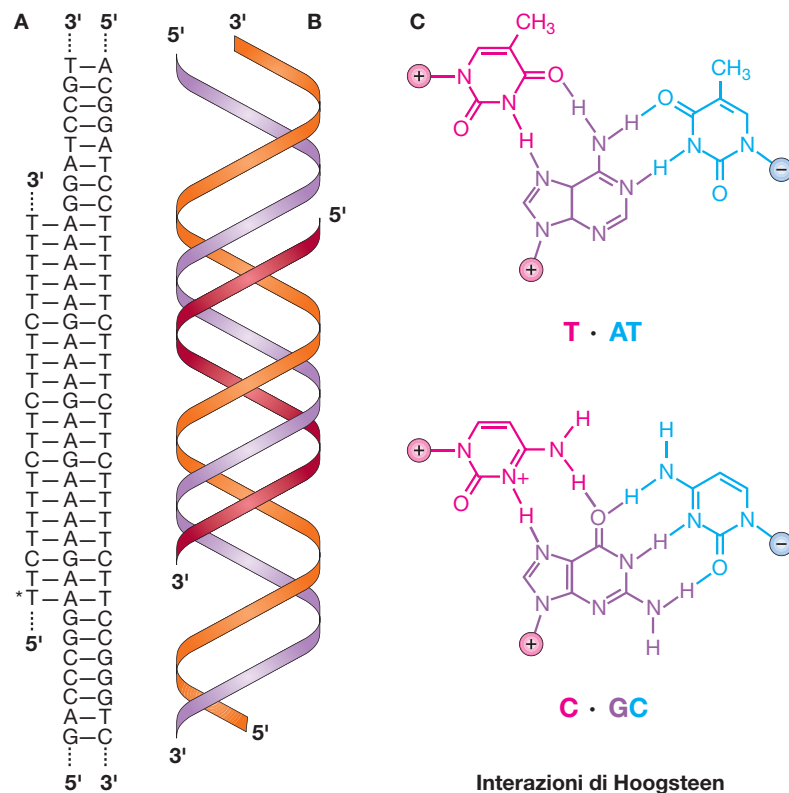


Figura 2.23 DNA a tripla elica. (A) Sequenze nucleotidiche dei due filamenti complementari del DNA e del terzo filamento che va a formare la struttura triplex; si notino le polarità 5' → 3' delle sequenze. (B) Rappresentazione schematica della struttura a tripla elica in cui si vede che il terzo filamento è inserito nel solco maggiore del duplex. (C) Appaiamenti tra le tre basi: si osservano i normali legami idrogeno del tipo Watson e Crick tra le basi complementari e gli appaiamenti di Hoogsteen con la terza base. In (B) e (C) è usato lo stesso codice colori.



2.22 illustra un tratto di DNA Z inserito tra due tratti di DNA B. È importante notare l'estruzione delle basi A e T nei punti di passaggio da una forma all'altra.

■ Tripla elica

In genere gli acidi nucleici, DNA e RNA, si trovano in forma di singolo filamento o di duplex, ma è stato poi scoperto che in casi molto particolari si possono formare anche delle regioni a tripla elica. Benché non sia stato dimostrato un ruolo naturale per questa struttura, essa riscuote un certo interesse perché potrebbe offrire la possibilità di sviluppare degli inibitori specifici per bloccare l'attività di geni bersaglio. Senza entrare nei dettagli, si può dire che, come mostrato in **figura 2.23**, un tratto

Figura 2.25 Sequenze ripetute e invertite e palindromi. (A) Sequenza palindromica, in cui le due sequenze ripetute sono adiacenti. (B) Sequenze ripetute e invertite separate tra loro da quattro pb.

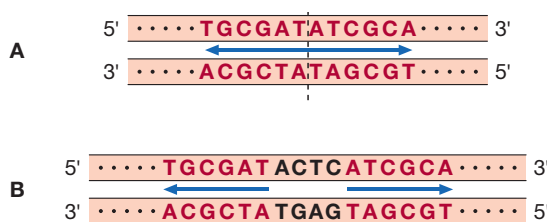
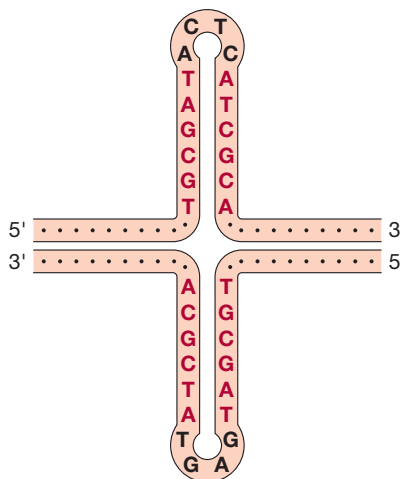


Figura 2.26 Struttura cruciforme. Un tratto di DNA a doppia elica contenente una sequenza ripetuta e invertita può formare degli appaiamenti intramolecolari e assumere una struttura cruciforme. La sequenza è la stessa mostrata in figura 2.25.



La situazione più comune è quella dei quartetti di G intramolecolari, in cui i tratti di DNA contenenti le serie di G si trovano tutti su un singolo filamento che, ripiegandosi tre volte su stesso, forma il quadruplex, come mostrato nella figura 2.24A (o anche, in forma diversa, in **fig. 2.24C I**). Esistono però varie altre situazioni: per esempio si possono avere quartetti di G intermolecolari in cui le quattro serie di G si possono trovare su due filamenti (**fig. 2.24C II**) o anche su quattro filamenti diversi (**fig. 2.24C III**). Si noti che, anche se in genere gli orientamenti $5' \rightarrow 3'$ dei quattro filamenti sono due a due antiparalleli, essi possono in alcuni casi anche avere tutti lo stesso orientamento, come è il caso del quadruplex intermolecolare a quattro filamenti mostrato nella figura 2.24C III. Quadruplex di quartetti di G si possono trovare *in vivo* in varie molecole strutturate di RNA, ma anche nel DNA. Per esempio è stato suggerito che, come vedremo più avanti (**par. 4.7 e fig. 4.13**), questa struttura si trovi nei telomeri dei cromosomi eucariotici.

■ Strutture cruciformi e strutture a forcina

Quando in un duplex di DNA (o anche di RNA) sono presenti, adiacenti o a breve distanza, due copie della stessa sequenza in orientamento opposto (**fig. 2.25**), queste vengono chiamate **ripetizioni invertite**. Se le due ripetizioni invertite sono adiacenti costituiscono, complessivamente, una **sequenza palindromica** che è una sequenza di DNA duplex identica su entrambi i filamenti se letti nella stessa direzione $5' \rightarrow 3'$. Le sequenze palindromiche, e più in genere le sequenze ripetute invertite, presentano interessanti proprietà strutturali e funzionali.

Dal punto di vista strutturale, un duplex che contenga una sequenza ripetuta e invertita può assumere una **struttura cruciforme**, come mostrato in **figura 2.26**. Infatti, ciascun filamento di una delle due sequenze ripetute trova una regione complementare cui appaiarsi adiacente nello stesso filamento. Le sequenze ripe-

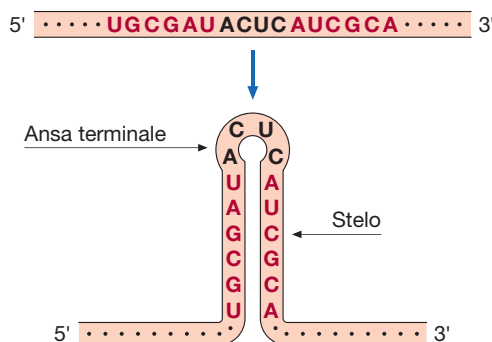


Figura 2.27 Struttura a forcina. Un tratto di RNA (o di DNA) costituito da un singolo filamento di una sequenza ripetuta e invertita a livello di doppia elica, può assumere una struttura a forcina (hairpin) costituita da uno stelo a doppia elica e da un'ansa terminale a singolo filamento.



Figura 2.28 Visualizzazione di una struttura cruciforme. Fotografia al microscopio elettronico di una molecola circolare di DNA contenente una struttura cruciforme. [Fonte: Wilson et al. (2001), *Journal of Molecular Biology*, 306(2), pp. 159-168.]

tute così appaiate intramolecolarmente costituiscono due “steli”, mentre le basi che separano le due sequenze ripetute non hanno modo di appaiarsi e formano due “anse”. Se prendiamo in considerazione un acido nucleico (DNA o RNA) a singolo filamento che abbia la stessa sequenza di uno dei due filamenti del duplex appena descritto, è evidente che potrà assumere una struttura con uno stelo a doppia elica e un'ansa. In questo caso, però, non abbiamo una struttura cruciforme, ma una **struttura a forcina** (fig. 2.27). La **figura 2.28** mostra un'immagine al microscopio elettronico di un DNA cruciforme.

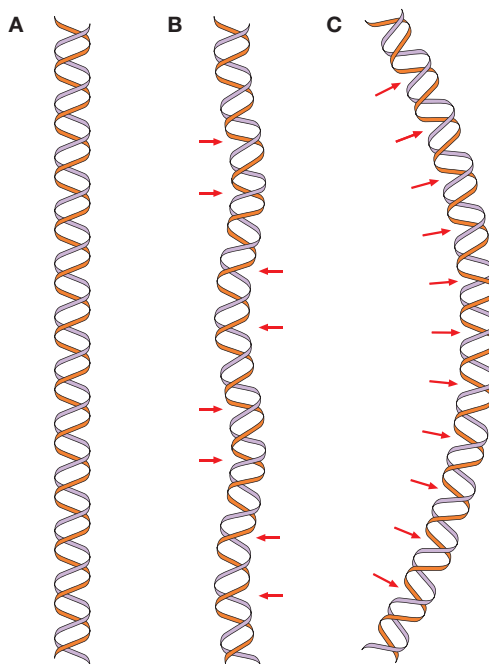
Come vedremo in dettaglio più avanti, queste strutture sono estremamente importanti soprattutto per l'RNA. Infatti, il DNA si trova in genere in forma di duplex in cui i due filamenti sono perfettamente complementari e quindi appaiati tra loro. L'RNA invece è solitamente costituito di un singolo filamento, prodotto della trascrizione di uno dei due filamenti di DNA. In questo caso è facile che brevi tratti di sequenza, perlopiù di pochi nucleotidi ma alle volte anche abbastanza estesi, possano trovare nelle adiacenze sequenze complementari con cui appaiarsi per formare delle forcine. Quindi, non ci si deve immaginare le molecole di RNA come singoli filamenti lineari distesi, ma piuttosto come strutture tridimensionali complesse comprendenti forcine a doppio filamento e regioni non appaiate. Solo per anticipare qui un esempio che poi riprenderemo più avanti, pensiamo alla struttura a trifoglio dei tRNA (RNA transfer; vedi fig. 2.51).

Poiché le basi sono solo di quattro tipi mentre le molecole di acidi nucleici sono spesso molto lunghe, è abbastanza facile che si trovino lungo la sequenza brevi tratti ripetuti e invertiti, e spesso si hanno vari modi alternativi di formare appaiamenti intramolecolari.

■ DNA curvo

I parametri della struttura B del DNA a doppia elica, come presentati da Crick e Watson, descrivono una doppia elica molto regolare. Nella realtà dei fatti, a causa delle interazioni di impilamento lungo la molecola del DNA ci sono variazioni strutturali locali che in alcuni casi possono influenzare la struttura complessiva molto regolare della molecola. Un caso ben studiato è quello del DNA intrinsecamente “curvo” o “piegato”. Se consideriamo due coppie di basi che si susseguono nel duplex, spesso queste non sono perfettamente parallele tra loro ma formano piccoli angoli, diversi a seconda delle particolari paia di basi considerate. Per esempio, due paia di basi A-T adiacenti hanno un'intrinseca tendenza a piegarsi dalla parte del solco minore, mentre due paia G-C hanno una tendenza inversa. Ne risulta che la doppia elica non è perfettamente dritta, ma presenta piccoli piegamenti (**fig. 2.29**). Questi, dato che la sequenza di nucleotidi è eterogenea e casuale, tendono a elidersi a vicenda, cosicché nell'insieme la struttura della molecola, pur con qualche deformazione, resta più o meno dritta (**fig. 2.29B**). Se però un tratto di DNA contiene, per esempio, due o tre paia di basi A-T consecutive, con una periodicità di circa 10 paia di basi (cioè ripetute a ogni giro di elica), i piccoli angoli introdotti si sommeranno tra loro producendo un'apprezzabile curvatura dell'asse della molecola di DNA (**fig. 2.29C**). Gli angoli che si formano tra coppie di basi adiacenti sono dovuti alle interazioni di impilamento che, come abbiamo già visto, sono descritte da quei parametri geometrici che abbiamo chiamato Tilt e Roll (**fig. 2.30**). L'accumulo di modificazioni sia del Tilt che del Roll nel DNA curvo possono risultare in strutture molto rigide, come nel caso del DNA del cinetoplasto del tripanosoma (**fig. 2.31**). La sequenza del DNA, quindi, può influenzare notevolmente la struttura della doppia elica e il fatto che queste strutture sono presenti *in vivo* suggerisce che la curvatura intrinseca di alcuni tratti del DNA abbia una funzione biologica. Più avanti, quando si descriverà la struttura della cromatina, parleremo ancora del DNA curvo e delle sue funzioni.

Figura 2.29 DNA intrinsecamente curvo (o piegato). (A) Struttura regolare di una molecola duplex di DNA secondo il modello di Crick e Watson. (B) Alcune coppie di basi adiacenti (per es. A-T/A-T), non essendo perfettamente parallele, introducono piccoli piegamenti della molecola duplex; se tali perturbazioni sono distribuite in maniera casuale lungo la molecola di DNA, risultano in una struttura un po' irregolare ma sostanzialmente dritta. (C) Se la sequenza nucleotidica è tale per cui i piccoli angoli tra le coppie di basi adiacenti si ripetono regolarmente a ogni giro di elica (cioè ogni 10 pb), gli angoli si sommano tra loro e risultano in una curvatura intrinseca della molecola di DNA.



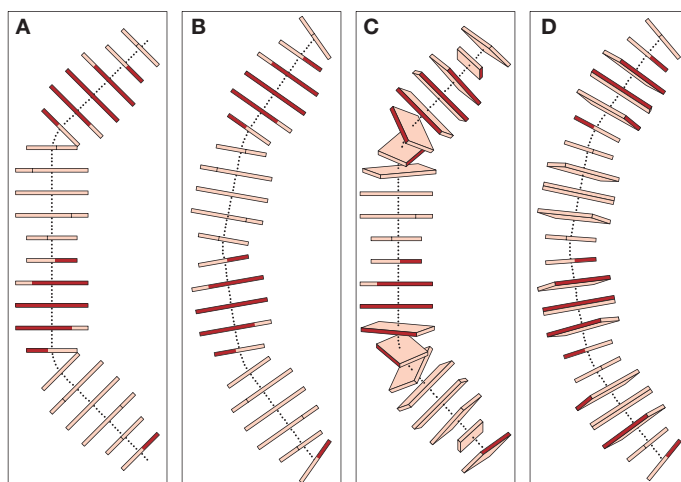


Figura 2.30 Il Tilt e il Roll delle coppie adiacenti di paia di basi sono all'origine di curvature dell'asse della doppia elica di DNA. Le modificazioni di Tilt e Roll possono essere più o meno drastiche e questo può causare diversi tipi di DNA piegato. In (A) il DNA è curvo per due forti piegamenti di Tilt adiacenti a un tratto regolare con le basi parallele. In (B) e (D) perturbazioni minori ma regolari generano un tratto di DNA piegato. (C) Due perturbazioni di Roll adiacenti a un tratto regolare formano una piegatura sul DNA.

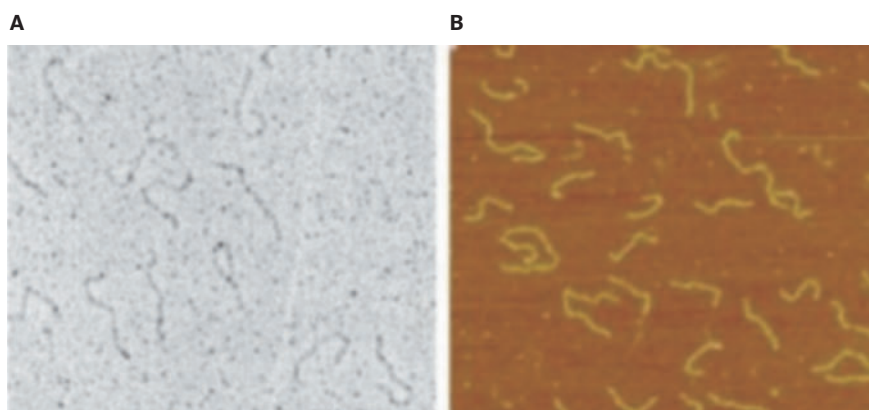
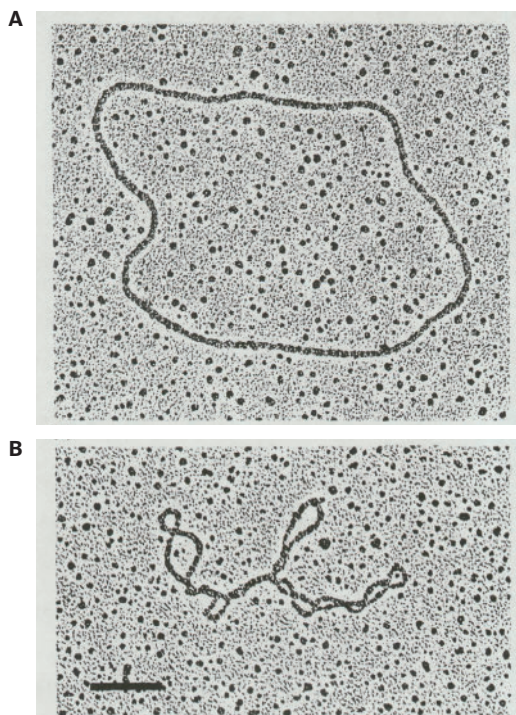


Figura 2.31 Tratto di DNA curvo del cinetoplasto del tripanosoma *Crithidia fasciculata*. La curvatura è più facilmente osservabile in frammenti piccoli di DNA. (A) Microscopia elettronica a trasmissione. (B) Microscopia a forza atomica. (Per gentile concessione di Gioacchino Micheli – Istituto di Biologia e Patologia Molecolari, CNR Roma.)

2.4 Topologia del DNA e DNA topoisomerasi

La struttura a doppia elica del DNA presenta molti vantaggi, per esempio una stabile e protetta conservazione dell'informazione genetica al suo interno, e grazie alla complementarietà dei due filamenti è possibile individuare e correggere errori e rotture che possono prodursi nel genoma. Tuttavia la struttura di due filamenti avvolti in una doppia elica pone dei seri problemi durante i vari processi che richiedono un'apertura dell'elica e la separazione dei due filamenti, come nella trascrizione e nella replicazione, dove il DNA si attorciglia su se stesso a formare strutture complesse, dette **superavvolgimenti**. Queste strutture "topologiche", se non controllate, possono interferire con l'espressione dei geni, la duplicazione del DNA e la segregazione dei cromosomi. Lo stato superavvolto del DNA contiene, come in una molla, energia che è utilizzata proprio per aprire i due filamenti o alle origini di replicazione o nelle regioni dei promotori e, quindi, se da un lato la cellula si sforza di preservare l'informazione contenuta nel genoma, dall'altro investe una considerevole quantità d'energia per mantenere lo stato topologico sotto controllo. Queste alterazioni topologiche avvengono sul DNA circolare covalentemente chiuso, come nel caso dei plasmidi o, come vedremo in seguito, su tratti di DNA lineari ma le cui estremità siano ancorate a strutture del cromosoma. Per controllare il grado di superavvolgimento del DNA che, se eccessivo, può essere dannoso, è necessario che il DNA venga tagliato, su

Figura 2.32 DNA circolare rilassato e superavvolto. Fotografie al microscopio elettronico (A) di una molecola circolare “rilassata” di DNA plasmidico e (B) di una molecola superavvolta. (Fonte: <http://www.cbs.dtu.dk>.)



uno o entrambi i filamenti, fatto ruotare e infine rilegato. Per svolgere al meglio questo compito, tutti gli organismi hanno selezionato una classe di enzimi che si chiamano **topoisomerasi** e che hanno la caratteristica di tagliare in maniera transiente uno o due filamenti del DNA.

Vediamo ora più in dettaglio quali sono le caratteristiche del DNA superavvolto e le regole matematiche che lo descrivono. Consideriamo due filamenti circolari chiusi che si avvolgono l'uno intorno all'altro nello spazio e immaginiamo che i due filamenti possano passare l'uno attraverso l'altro ed essere così separati. Il numero di volte che un filamento dovrebbe passare attraverso l'altro, in maniera che essi possano essere completamente separati generando due cerchi a singolo filamento, si chiama **numero di legame**. La frequenza (quante volte un filamento si avvolge sull'altro) e la posizione di entrambi nello spazio tridimensionale possono essere descritte da due grandezze: la prima è il **twist** (Tw , “torsione”), la seconda è il **writhe** (Wr , “contorsione”). Considerando il DNA in generale, il **twist** rappresenta il numero di giri della doppia elica rispetto all'asse centrale e definisce così il grado di avvitamento della doppia elica. In un DNA superavvolto, il **writhe** è il numero di volte che l'asse centrale della doppia elica incontra se stesso, formando i superavvolgimenti che sono visibili per esempio al microscopio elettronico (la **figura 2.32** mostra una fotografia di un DNA circolare rilassato e superavvolto). La somma di queste due grandezze, twist e writhe, indica il numero di volte che un filamento si avvolge sull'altro ed è appunto il già ricordato numero di legame (Lk , *linking number*):

$$Lk = Tw + Wr$$

Questa equazione, che fa parte di quel settore della matematica che si chiama topologia, descrive le possibili conformazioni che il DNA assume nello spazio tridimensionale.

In un DNA circolare covalentemente chiuso, il numero di legame, che è sempre un numero intero, corrisponde al numero totale delle coppie di basi N diviso per

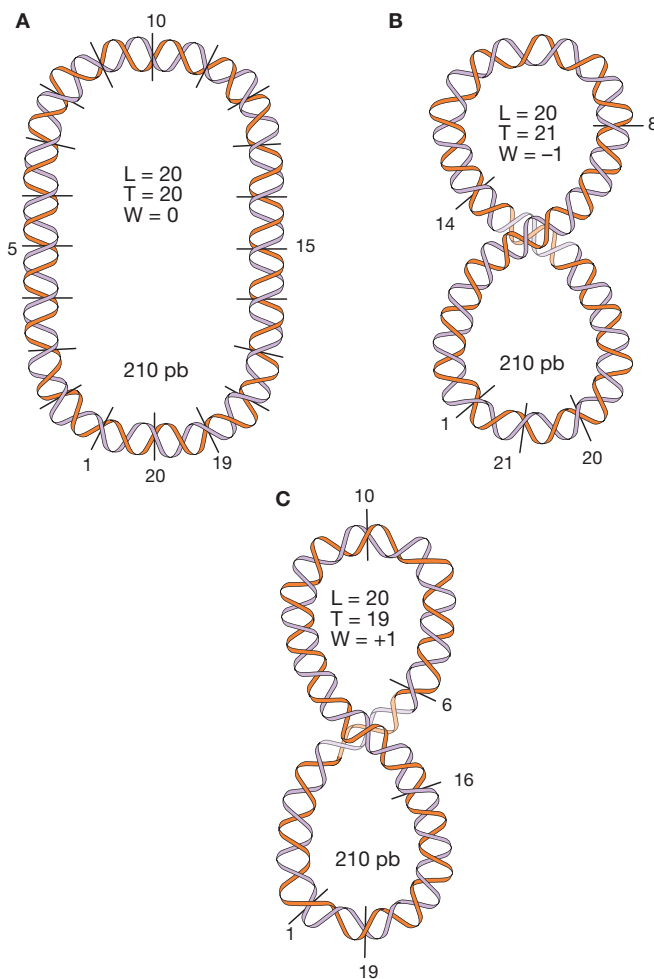


Figura 2.33 Variazioni del twist e del writhe in una molecola di DNA circolare covalentemente chiusa, quindi con numero di legame costante. Se si forza il DNA circolare a cambiare il suo writhe, senza tagliare la doppia elica, si avrà un compensativo cambiamento del twist, mentre il numero di legame rimarrà lo stesso. (A) Molecola di DNA circolare rilassata. (B) In una molecola in cui è stato introdotto un giro di superavvolgimento negativo (writhe = -1) il twist è positivo passando da 20 a 21. (C) In una molecola in cui è stato introdotto un giro di superavvolgimento positivo (writhe = +1) il twist è negativo passando da 20 a 19.

10,5, che è il numero di basi per un giro d'elica. Per un DNA lungo 210 coppie di basi il numero di legame è 20. Secondo l'equazione descritta sopra, se si forza il DNA circolare a cambiare il suo writhe, come per esempio in **figura 2.33**, senza tagliare la doppia elica, per compenso cambierà il twist, ma il numero di legame sarà lo stesso. In B il writhe è negativo e quindi il twist è positivo, in C il writhe è positivo e quindi il twist sarà negativo. Per alterare il numero di legame bisogna rompere la doppia elica del DNA e far ruotare i due filamenti uno rispetto all'altro per aumentare o ridurre il numero di volte che un filamento si incrocia con l'altro. Se, per esempio, dopo aver aperto il cerchio (**fig. 2.34**) si riduce il numero di legame di 2 unità e lo si richiude, la molecola tenderà a ritrovare il suo equilibrio naturale di DNA B con 10,5 coppie di basi per giro e quindi compenserà la riduzione per riportare il twist a 20 con l'introduzione di writhe negativo, assumendo una conformazione con 2 superavvolgimenti. Al contrario, se si taglia il DNA e si aumenta il numero di legame di 2 unità agendo sul twist, una volta che la molecola è richiusa il twist tenderà a tornare a 20 con la formazione di 2 superavvolgimenti (writhe) positivi. Si può compensare la diminuzione del numero di legame anche denaturando il DNA, e questo avviene spesso in regioni ricche in AT. In natura il DNA tende ad avere un L_k minore rispetto alla sua lunghezza e ciò permette di avere una riserva di energia contenuta nella struttura "a molla" dei superavvolgimenti negativi, che facilita, per esempio, l'apertura della doppia elica all'inizio della replicazione o della trascrizione. In termini quantitativi,

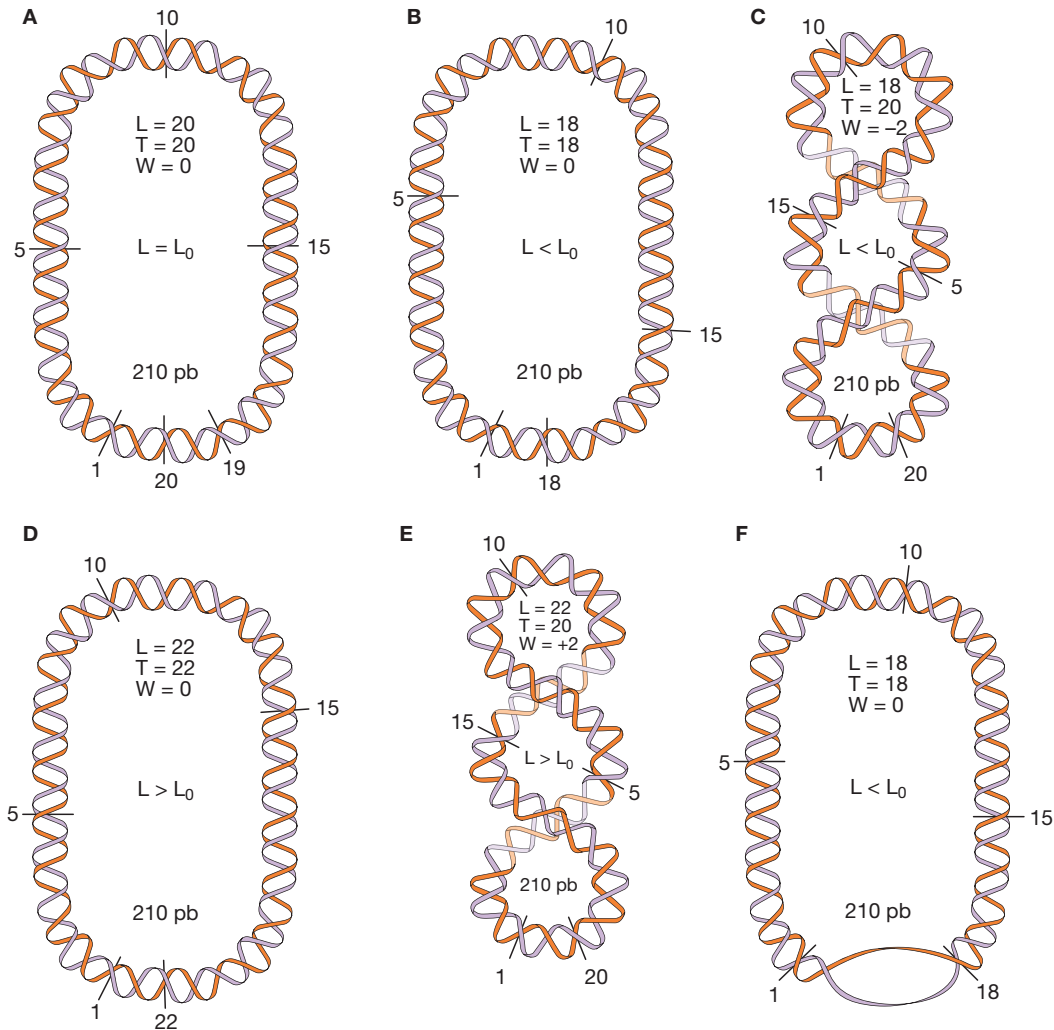


Figura 2.34 Variazione del numero di legame. Il numero di legame di una molecola di DNA circolare può essere variato tagliando uno o ambedue i filamenti del duplex, introducendo uno o più rotazioni a una estremità e richiudendo covalentemente i due filamenti della molecola circolare. L_0 indica il numero di legame che compete alla molecola di DNA di questa lunghezza. (A) Molecola di DNA circolare rilassata covalentemente chiusa. (B) La molecola di DNA circolare è stata aperta, il numero di legame ridotto di 2 unità e poi richiusa; si può notare la variazione del numero di legame. (C) La molecola tende a ritrovare il suo equilibrio naturale di DNA B

con 10,5 coppie di basi per giro; compenserà quindi la riduzione del twist riportandolo a 20, con l'introduzione di writhe negativo e assumendo una conformazione con 2 superavvolgimenti negativi. (D) La molecola di DNA circolare è stata aperta e il numero di legame aumentato di 2 unità agendo sul twist. (E) Quando la molecola viene richiusa, il twist tenderà a tornare a 20 con la formazione di 2 superavvolgimenti (writhe) positivi. (F) Si può compensare la diminuzione del numero di legame anche denaturando il DNA e questo avviene spesso in regioni ricche in AT. La numerazione a fianco delle molecole si riferisce al numero di giri di doppia elica.

l'energia contenuta in un plasmide circolare di circa 5 kb che si trova nel suo stato naturale negativamente superavvolto è di circa $50\text{-}100 \text{ kcal mole}^{-1}$.

Esistono due tipi di superavvolgimenti: quello che vediamo nel DNA nudo, come nelle immagini al microscopio elettronico, si chiama **avvolgimento plectonemico**; quello che invece si ha quando il DNA è avvolto intorno all'ottamero istonico si chiama **avvolgimento toroidale** e da un punto di vista topologico è equivalente. Esistono anche molecole circolari annodate che sono generalmente il prodotto di eventi di ricombinazione. Queste strutture topologiche sono variazioni del writhe. Molecole di DNA circolari covalentemente chiuse di uguale lunghezza, ma che differiscono solo per il numero di legame, sono definite **topoisomeri**. I topoisomeri che differiscono anche per un solo numero di legame o nodi con vari incroci possono essere risolti mediante elettroforesi su gel di agarosio (fig. 2.35).

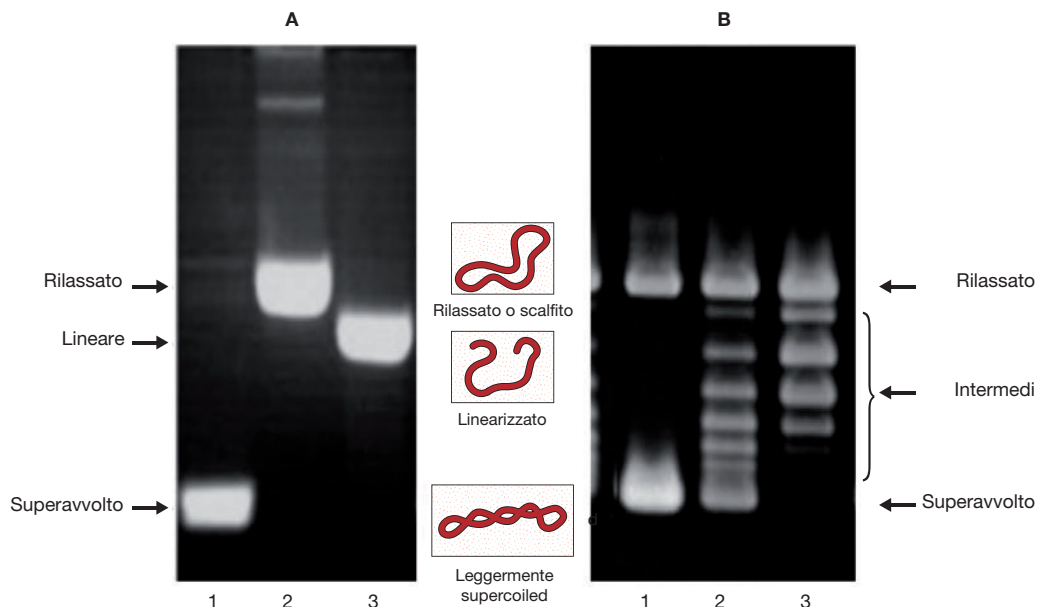


Figura 2.35 Frazionamento di topoisomeri mediante elettroforesi su gel di agarosio. (A) Elettroforesi di tre campioni della stessa molecola di DNA plasmidico nelle forme circolare superavvolta (corsia 1), circolare rilassata mediante taglio di uno dei due filamenti del duplex (corsia 2) e linearizzata mediante taglio con una endonucleasi (corsia 3); le tre forme topoisomeriche sono mostrate

schematicamente a destra del gel. (B) Una preparazione di DNA plasmidico nella forma superavvolta (corsia 1), che contiene anche una certa quantità di forma rilassata, è stata trattata per tempi crescenti (corsie 2 e 3) con una topoisomerasi che diminuisce in maniera controllata il superavvolgimento, producendo intermedi con numero di giri di superavvolgimento progressivamente minori.

Ci sono dei reagenti che si intercalano nel DNA che possono cambiarne la topologia. Il bromuro di etidio (**fig. 2.36**), che viene normalmente usato per visualizzare il DNA perché è fluorescente se esposto ai raggi ultravioletti, è una molecola planare che si intercala tra le basi. Gli effetti dell'intercalazione sono molteplici: in primo luogo la molecola di DNA si allunga perché due coppie di basi adiacenti devono allontanarsi per far posto alla molecola di etidio; inoltre una volta che l'etidio è all'interno della doppia elica, per le interazioni di impilamento descritte precedentemente, diminuisce il twist di 26° . Sono necessarie quindi 13,8 molecole di etidio per srotolare di 360° la doppia elica e togliere un giro completo di superavvolgimento. È importante notare che l'etidio si lega meglio a un DNA negativamente superavvolto.

Anche i lunghi cromosomi lineari degli eucarioti si comportano come molecole circolari, dal momento che sono organizzati in grandi anse ancorate alla matrice nucleare, e quindi anche in questo caso il DNA, che è avvolto intorno agli istoni, obbedisce alle regole della topologia descritte (**fig. 2.37**).

Consideriamo ora il caso di un DNA circolare completamente rilassato, cccDNA (*covalently closed circular DNA*), il suo numero di legame Lk sarà uguale al suo twist e il writhing sarà 0. Definiamo in questo caso $Lk = Lk_0$. Se vogliamo paragonare il grado di superelicità di due DNA che hanno la stessa lunghezza, come due diversi topoisomeri, questa differenza è definita da

$$\Delta Lk = Lk - Lk_0$$

Se il ΔLk di un cccDNA è diverso da 0, la molecola conterrà della tensione e sarà superavvolta. Con $Lk < Lk_0$ e $\Delta Lk < 0$, il DNA è superavvolto negativamente, se $Lk > Lk_0$ e $\Delta Lk > 0$, allora il DNA è superavvolto positivamente.

Per paragonare il grado di superelicità di DNA di grandezza diversa, è più comodo usare, invece del ΔLk , la densità di superelica, che ha come simbolo σ ed è definita da $\sigma = \Delta Lk / Lk_0$.

Figura 2.36 Bromuro di etidio, un intercalante che modifica la topologia del DNA. Il bromuro di etidio è una molecola idrofobica planare che si intercala tra le coppie di basi della doppia elica del DNA cambiandone in vari modi la topologia: la doppia elica si allunga, il twist diminuisce e diminuisce il superavvolgimento.

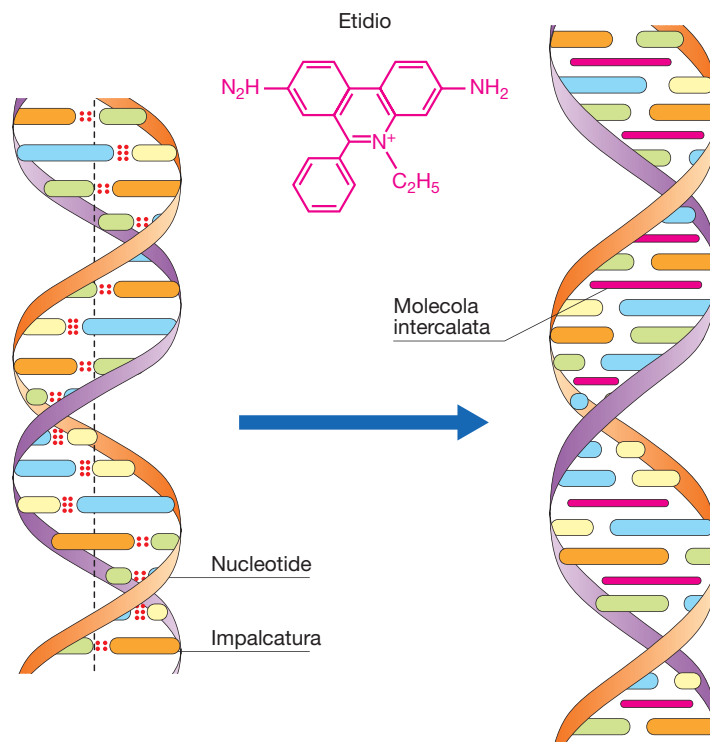
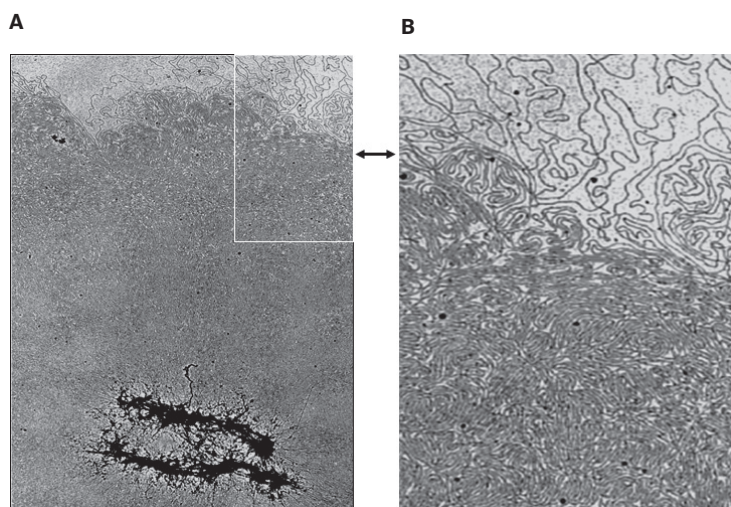


Figura 2.37 Visualizzazione al microscopio elettronico di un cromosoma eucariotico privato degli istoni. (A) Le lunghissime molecole di DNA cromosomico sono organizzate in numerose anse ancorate alla base a una residua impalcatura proteica che ancora mantiene la struttura del cromosoma. Ciascuna ansa, essendo ancorata alla base, si comporta, dal punto di vista topologico, come una molecola circolare, potendo assumere giri di superavvolgimento indipendentemente dalle anse adiacenti. (B) Ingrandimento di una regione periferica della foto (A) in cui si apprezza come intorno al cromosoma ci sia una fitta matassa di DNA che solo in periferia si dirada permettendo di vedere il filamento del DNA. (Fonte: Ulrick Laemmli.)



In natura le molecole di DNA circolare che si trovano in cellule procariotiche ed eucariotiche sono in genere superavvolte negativamente con un valore di σ intorno a $-0,05$, $-0,06$.

Come abbiamo detto precedentemente, il superavvolgimento contiene energia libera perché il DNA tende a ritornare nel suo stato di quiete e, quindi, il DNA negativamente superavvolto faciliterà quei fenomeni biologici dove i due filamenti devono essere parzialmente denaturati.

Una molecola di DNA duplex chiusa in cui introduciamo giri di superavvolgimento negativi tende, fino a che può, a scaricare questa tensione torsionale (energia libera) arrotondandosi in senso opposto. La molecola così superavvolta contiene un eccesso di energia libera che può permettere al DNA di subire delle transizioni

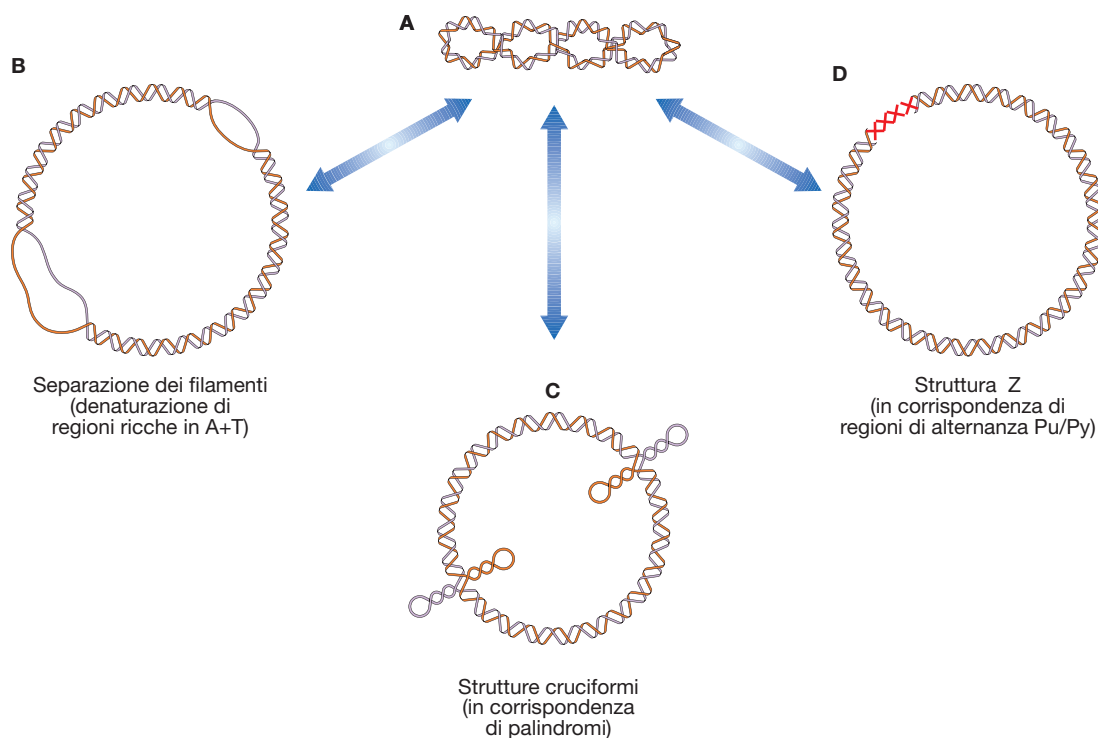


Figura 2.38 L'energia libera contenuta in una molecola superavvolta può indurre transizioni strutturali del DNA. Una molecola circolare covalentemente chiusa e superavvolta negativamente (A) può scaricare la tensione torsionale (energia libera) in varie maniere: il duplex si può denaturare localmente, preferenzialmente

in regioni ricche in A-T (B); se sono presenti sequenze ripetute e invertite, si può avere la formazione di strutture cruciformi (C); se sono presenti tratti con alternanza purina/pirimidina il DNA può assumere localmente la struttura sinistrorsa Z (D).

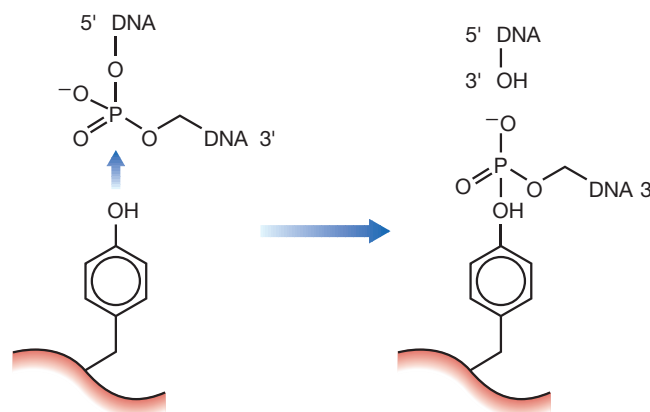
strutturali che non sarebbero possibili nel DNA rilassato. Una di queste transizioni consiste nella separazione (denaturazione) dei due filamenti per un breve tratto (**fig. 2.38B**). Ogni giro di superelica può essere “riassorbito” dalla denaturazione di una sequenza di circa dieci paia di basi. È evidente che questo avviene con più facilità in tratti di DNA ricchi in paia di basi A-T. Un'altra transizione che può essere favorita dalla presenza di un eccesso di energia in una molecola di DNA superavvolta, è la formazione di strutture cruciformi (**fig. 2.38C**). Infatti, se nella molecola esiste una sequenza ripetuta e invertita (palindromica), questa può assumere una struttura cruciforme sottraendo così giri di superavvolgimento dalla molecola nel suo insieme. Infine, un terzo tipo di cambiamento strutturale indotto da un superavvolgimento negativo è la transizione da forma B a forma Z di un tratto di DNA che abbia la propensione a questa transizione, cioè in una regione in cui ci sia un'alternanza regolare di pirimidine e purine (**fig. 2.38D**).

In natura il DNA positivamente superavvolto si trova in particolari organismi termofili che vivono a temperature molto elevate. La funzione biologica di questo stato topologico è quella di evitare che il DNA si denaturi con troppa facilità. Alcuni di questi organismi hanno delle DNA topoisomerasi in grado di introdurre superavvolgimenti positivi sul DNA.

DNA topoisomerasi

Lo stato topologico del DNA deve essere tenuto sotto controllo e per fare questo la doppia elica deve essere aperta e richiusa temporaneamente: vale a dire che uno o entrambi i filamenti del DNA devono essere tagliati, manipolati rispetto ai filamenti interi e poi rilegati. Per catalizzare queste reazioni complesse gli

Figura 2.39 Meccanismo di taglio e rilegazione delle topoisomerasi. La tirosina “attiva” della topoisomerasi fa un attacco nucleofilo al gruppo fosfato dello scheletro del DNA con il suo gruppo OH, formando un legame covalente tra l’enzima e il DNA. La reazione inversa di rilegazione è prodotta da un attacco nucleofilo del gruppo OH dell’estremità del DNA libera che rilega il fosfato, formando di nuovo la doppia elica e liberando l’enzima. [Fonte: Wang, J.C. (2002), *Nature Review Molecular Cell Biology*, 3, pp. 430-440.]



organismi contengono una classe di enzimi, chiamati **DNA topoisomerasi**, che hanno la caratteristica comune di creare un taglio transiente sul DNA, mediante una tirosina che si lega covalentemente allo scheletro fosfato, con una reazione di transesterificazione fatta dal suo gruppo OH. Successivamente, dopo la manipolazione topologica che riduce o aumenta il numero di legame, il gruppo OH libero del filamento tagliato, attraverso una reazione inversa, attacca il legame tra la tirosina e il DNA, ripristinando la continuità della doppia elica (**fig. 2.39**). Anche se formalmente la rottura e la rilegazione del legame fosfodiesterico del DNA non richiede energia, l'intero processo di modificazione topologica utilizza l'energia libera contenuta nel DNA superavvolto.

Le DNA topoisomerasi (a volte chiamate anche TOPO o TOP) si differenziano tra loro per il meccanismo d'azione con il quale cambiano la topologia del DNA, ma esistono due principali strategie per portare avanti queste reazioni. Nella prima, chiamata **rotazione controllata**, l'enzima introduce una singola rottura su un filamento della doppia elica e permette la rotazione su sé stesso del filamento intatto per un numero variabile di giri, fino a che l'attrito tra il DNA e l'enzima induce la rilegazione del filamento tagliato. Il secondo meccanismo, comune a molte DNA topoisomerasi, è definito **strand passage** (passaggio del filamento), e consiste nel creare una rottura singola o a doppio filamento sul DNA, allargare l'interruzione prodotta e far passare l'altro filamento o la doppia elica attraverso la rottura che successivamente verrà risaldata.

Il modo con il quale il DNA viene tagliato è alla base della classificazione di questi enzimi. Le **topoisomerasi di tipo I**, che sono in genere dei monomeri, tagliano un solo filamento del DNA al 5', creando in questo modo un'apertura mediata dalle interazioni di diversi domini dell'enzima con la doppia elica, attraverso la quale può passare l'altro filamento o una doppia elica. Questo meccanismo, che come abbiamo già ricordato è definito strand passage, permette di eliminare con grande efficienza strutture annodate sul DNA, nonché di decatenare due filamenti di cui uno contenga una rottura a singolo filamento e di rilassare esclusivamente superavvolgimenti negativi. La **figura 2.40A** mostra, come esempio, la struttura cristallografica della topoisomerasi IB; questa avvolge la doppia elica come una tenaglia, e le interazioni tra i vari domini dell'enzima con il DNA contribuiscono all'attrito necessario a controllare la rotazione del filamento intatto su se stesso. Il meccanismo d'azione della topoisomerasi IA è schematizzato in **figura 2.40B**.

Tutte le **topoisomerasi di tipo II**, che sono in genere dei dimeri o multimeri, introducono un taglio su entrambi i filamenti del DNA con le due tirosine covalentemente legate all'estremità 5' e portano avanti le modificazioni topologiche facendo passare un secondo tratto a doppia elica attraverso la rottura. Il filamento tagliato si chiama segmento G per *gate* (cancello), il segmento che passa attraverso

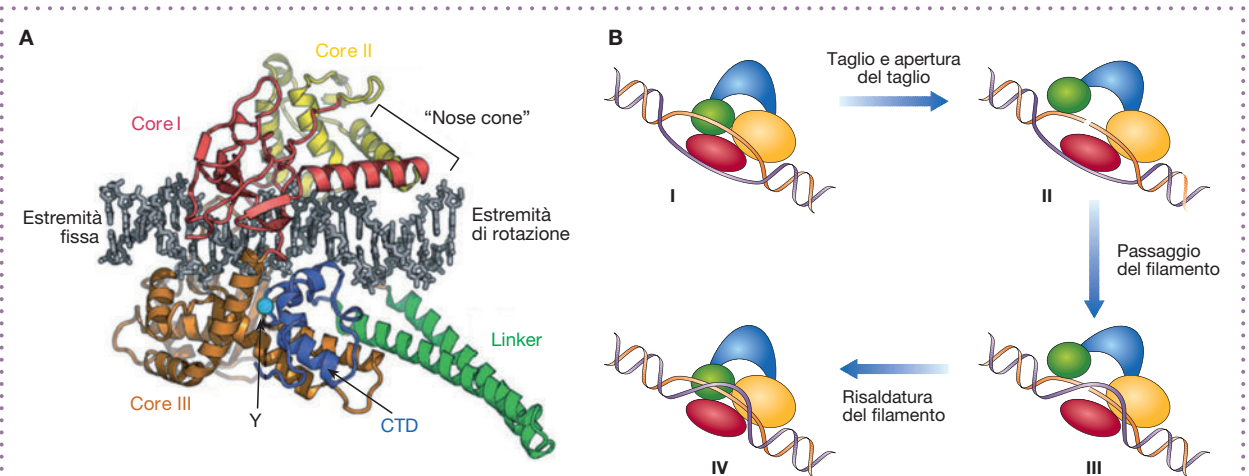


Figura 2.40 Struttura e meccanismo d'azione delle topoisomerasi di tipo I. (A) Struttura cristallografica della topoisomerasi IB, enzima monomero costituito da vari domini qui indicati con diversi colori; notare il posizionamento della doppia elica di DNA (grigio) che è libera

di ruotare. (B) Meccanismo d'azione della topoisomerasi IA: (I) l'enzima lega il DNA e apre la doppia elica; (II) uno dei due filamenti del DNA viene tagliato; (III) il filamento integro di DNA passa attraverso l'apertura del primo filamento; (IV) il filamento tagliato viene rilegato.

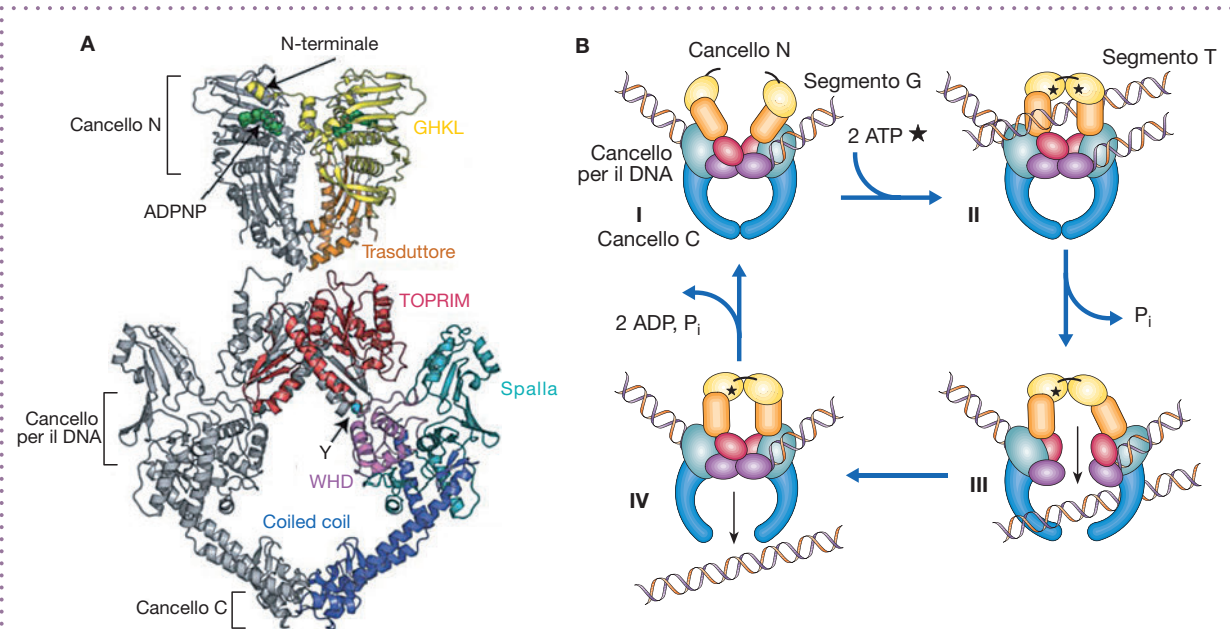


Figura 2.41 Struttura e meccanismo d'azione delle topoisomerasi di tipo II. (A) Struttura tridimensionale dell'enzima multimerico, costituito da più subunità qui indicate con diversi colori; si notino il cancello N e, in basso, il cancello per il DNA. Sono visibili anche i domini di legame dell'ATP (ADPNP). L'enzima è costituito da due domini: il dominio N-terminale (sopra) e il dominio C-terminale (sotto), che qui sono mostrati ambedue in conformazione chiusa e un po' distanziati tra loro. Nella struttura complessiva i due cancelli si aprono indipendentemente e alternativamente co-

me descritto in (B). (B) Modello del "doppio cancello". (I) L'enzima ha una conformazione aperta pronta ad accettare al suo interno il segmento T (trasporto), mentre ha già al suo interno il segmento di DNA G (cancello); (II) il legame con l'ATP genera un cambio di conformazione e il cancello N si chiude; successivamente l'idrolisi dell'ATP genera un nuovo cambio di conformazione; (III) il DNA G viene tagliato e il DNA T passa attraverso il taglio provocando l'apertura del secondo cancello; (IV) il DNA T viene espulso dall'enzima.

so l'apertura si chiama segmento T per *transport*. La **figura 2.41** illustra la struttura e il meccanismo di azione delle topoisomerasi di tipo II, che viene definito **a doppio cancello**. In questo processo il segmento T del DNA entra nella parte superiore dell'enzima, che si apre per accoglierlo; dopo l'apertura della doppia elica del segmento G, il segmento di T attraversa tutto l'enzima, che si apre nella parte inferiore, rimuovendo in questa maniera due superavvolgimenti alla volta.

FINESTRA 2.1

VARIETÀ E CLASSIFICAZIONE DELLE DNA TOPOISOMERASI

In tutte le cellule, procariotiche ed eucariotiche, sono presenti vari tipi di topoisomerasi, di **tipo I** e di **tipo II**, coinvolte in diversi processi cellulari quali replicazione, trascrizione, ricombinazione ecc. La **figura 2F1.1** presenta le DNA topoisomerasi conosciute e le loro funzioni nel controllare la topologia del DNA.

Come descritto nel testo (e in fig. 2.40), gli enzimi di tipo I tagliano un solo filamento del DNA e vi si legano all'estremità 3' creata; essi sono in grado di rilassare sia superavvolgimenti positivi che negativi mediante la "rotazione controllata" del filamento intatto su se stesso. Le topoisomerasi di tipo I si dividono in tre sottofamiglie: IA, IB, IC. Le topoisomerasi di **tipo IB** includono la topoisomerasi I eucariotica, quella dei poxvirus e di alcuni batteri. Alla famiglia **IC** appartiene la topoisomerasi V presente in alcune specie di archeobatteri e con un meccanismo per certi aspetti simile alla topoisomerasi IB. Le topoisomerasi di **tipo IA** includono le topoisomerasi I e III dei batteri e degli archeobatteri e la topoisomerasi III degli eucarioti. A questo gruppo appartiene anche un enzima degli archeobatteri termofili, la "girasi inversa", che è in grado di introdurre superavvolgimenti positivi sul DNA. Fisiologicamente questa caratteristica aiuta a contrastare la tendenza del DNA a essere denaturato alle alte temperature. Tutti questi enzimi richiedono per funzionare la presenza di ioni Mg^{++} o Zn^{++} e, inoltre, la girasi inversa necessita di ATP.

Le DNA topoisomerasi di tipo II, il cui meccanismo d'azione è descritto nel testo (e in fig. 2.41), sono sottoclassificate come IIA o IIB in base alla loro struttura e alla loro sequenza amminocidica. A loro volta, le **topoisomerasi IIA** comprendono:

1. la topoisomerasi II eucariotica;
2. le topoisomerasi dei batteriofagi;
3. la girasi dei batteri, degli archea e dei cloroplasti;
4. la topoisomerasi IV dei batteri.

Le topoisomerasi II e IV possono rilassare superavvolgimenti positivi e negativi, la girasi è in grado di rimuovere con efficienza solo superavvolgimenti positivi ed è l'unica topoisomerasi in grado di introdurre superavvolgimenti negativi. Tutte le topoisomerasi IIA sono in grado di decatenare e rimuovere nodi sul DNA e la topoisomerasi IV è la più efficiente in quest'ultima reazione. Le **topoisomerasi IIB** sono confinate al mondo degli archea. La topoisomerasi VI è la rappresentante maggiore di questo gruppo e il suo meccanismo d'azione è molto simile a quello delle topoisomerasi IIA.

La distribuzione filogenetica delle topoisomerasi è complessa, ma una caratteristica comune a tutti gli organismi cellulari è quella di avere almeno una DNA topoisomerasi di tipo I e una di tipo II. Nei batteri come *Escherichia coli*, per esempio, la girasi mantiene alto il livello dei superavvolgimenti negativi nel cromosoma circolare per il suo compattamento nella cellula e poiché questo batterio si divide molto rapidamente, la decatenazione dei cromosomi alla fine della replicazione è garantita dalla topoisomerasi IV, che è molto efficiente in questa reazione. Inoltre, il controllo omeostatico del livello dei superavvolgimenti del batterio è garantito dall'azione coordinata della girasi, che introduce superavvolgimenti negativi, e della topoisomerasi IA, che invece è in grado di rimuovere questo tipo di superavvolgimenti. Generalmente, se avvengono delle mutazioni nel gene

della topoisomerasi IA, il batterio, per sopravvivere, acquisisce mutazioni compensatorie nei geni della girasi. Sempre in *E. coli*, i problemi di ricombinazione che possono avvenire durante la replicazione sono controllati dalla topoisomerasi III. Negli eucarioti esistono varie combinazioni nella presenza delle diverse topoisomerasi. Qui, infatti, la presenza degli istoni confina il ruolo di questi enzimi a tenere sotto controllo la struttura topologica del DNA durante fenomeni come la replicazione, la trascrizione e la ricombinazione nel contesto della struttura della cromatina. La topoisomerasi IB è essenziale durante la trascrizione e la IIA per decatenare il DNA alla fine della replicazione, mentre la III risolve strutture ricombinanti che possono formarsi durante la replicazione.

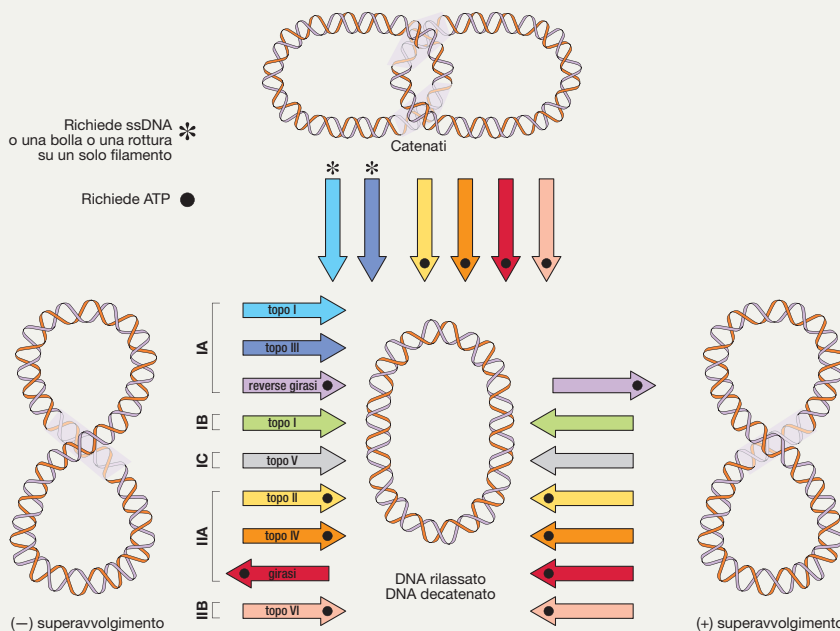


Figura 2F1.1 Topoisomerasi, reazioni di rilassamento, catenazione e decatenazione del DNA, introduzione di superavvolgimenti negativi o positivi. Le frecce indicano la direzione della reazione. (Per gentile concessione di James Berger, Università di California, Berkeley.)

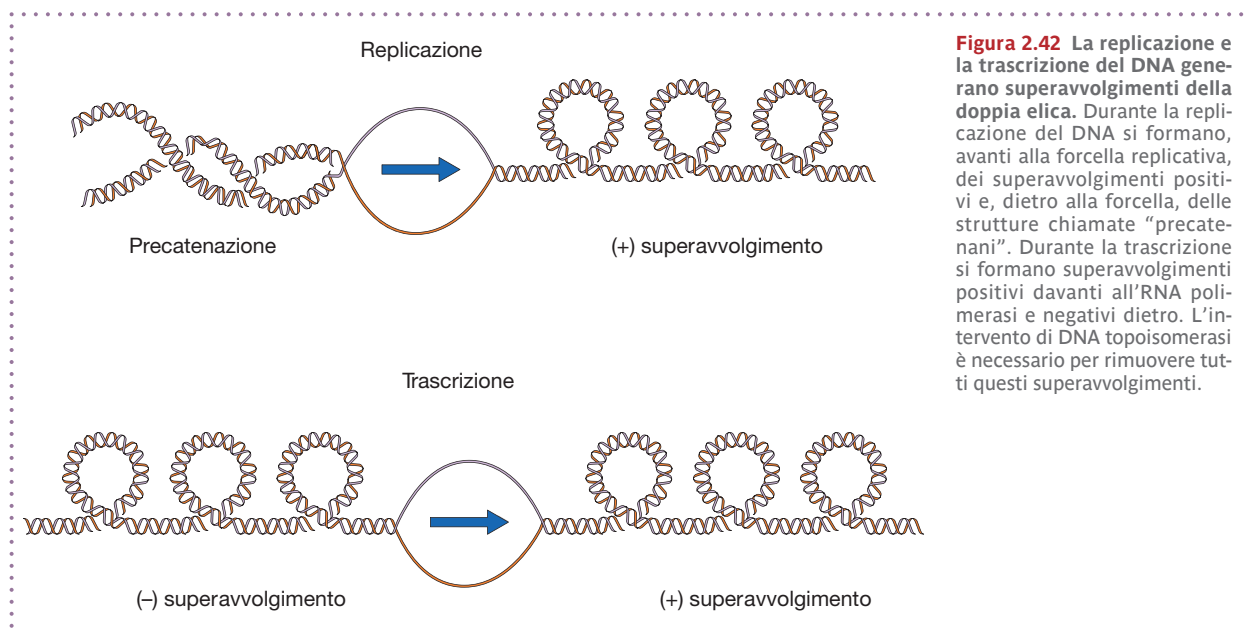


Figura 2.42 La replicazione e la trascrizione del DNA generano superavvolgimenti della doppia elica. Durante la replicazione del DNA si formano, avanti alla forcella replicativa, dei superavvolgimenti positivi e, dietro alla forcella, delle strutture chiamate "precatenani". Durante la trascrizione si formano superavvolgimenti positivi davanti all'RNA polimerasi e negativi dietro. L'intervento di DNA topoisomerasi è necessario per rimuovere tutti questi superavvolgimenti.

L'idrolisi dell'ATP è necessaria per promuovere i cambi conformazionali che permettono al DNA di entrare e passare attraverso l'enzima.

Come approfondito nella **finestra 2.1**, per ciascun tipo di topoisomerasi ne esistono, sia nei procarioti che negli eucarioti, molte sottoclassi che si distinguono sia per la sequenza amminoacidica che per il meccanismo d'azione.

Tutti quei processi cellulari che consistono nell'apertura della doppia elica e nello scorrimento lungo il DNA di grossi apparati, come il complesso di replicazione o trascrizione, generano superavvolgimenti sulla doppia elica (**fig. 2.42**). Infatti, è stato dimostrato che *in vivo* per la viscosità del nucleo o nel citoplasma dei batteri questi grossi apparati multiproteici non riescono a ruotare intorno al DNA. Inoltre il DNA stesso non è libero di ruotare facilmente per varie ragioni, tra le quali la sua lunghezza e la presenza di siti di ancoraggio a strutture cellulari come la membrana nucleare. Questa condizione genera superavvolgimenti positivi nella direzione della sintesi davanti al complesso che avanza e superavvolgimenti negativi a monte. Queste regioni di DNA possono diventare superavvolte fino al punto che l'apparato non è più in grado di procedere: le DNA topoisomerasi sono quindi essenziali per tenere sotto controllo il grado di superavvolgimento del DNA durante questi processi.

2.5 Struttura dell'RNA

Il ruolo e l'importanza dell'RNA sono stati presi in considerazione immediatamente dopo la scoperta della doppia elica del DNA. Infatti, in un rapporto inviato nel 1955 ai membri del Club della cravatta a RNA (RNA Tie Club) (**finestra 2.2**), Crick predisse l'esistenza di adattatori che permettessero il trasferimento dell'informazione da uno stampo di RNA alla sequenza di amminoacidi nelle proteine. Questa osservazione è sorprendentemente preveggenza dal momento che all'epoca non erano ancora stati isolati l'RNA messaggero e il tRNA, che si è rivelato successivamente essere proprio l'adattatore. Come si vedrà in seguito nel capitolo sulla traduzione (*vedi par. 14.2*), con la sua duplice natura il tRNA fa da tramite tra il linguaggio del DNA e dell'RNA e il linguaggio delle proteine. Successivamente, nel 1960 fu isolato l'RNA messaggero da S. Brenner, F. Jacob e M. Meselson. L'RNA ribosomale era noto da tempo.

FINESTRA 2.2

L'RNA TIE CLUB

Dopo la risoluzione della struttura a doppia elica del DNA, molti scienziati erano curiosi e ansiosi di capire come l'informazione in esso contenuta poteva essere decifrata. Il fisico russo George Gamow, che frequentava i laboratori di Cambridge dove lavoravano Watson e Crick, credeva che la maniera migliore per risolvere il mistero sarebbe stata quella di riunire tra loro scienziati di discipline diverse per mettere in comune idee e risultati. Nel 1954 fondò l'RNA Tie Club con l'obiettivo di risolvere l'enigma della struttura dell'RNA e capire come esso interviene nella costruzione delle proteine. Il Club consisteva di 20 membri regolari, uno per ciascun amminoacido, e 4 membri onorari, per le basi del DNA. Tutti i membri indossavano una cravatta disegnata dallo stesso Gamow che rappresentava un'ipotetica molecola di RNA (fig. 2F2.1). C'erano molti scienziati importanti, otto dei quali avevano o avrebbero ricevuto il premio Nobel. Watson aveva il codice Prolina, Crick Tirosina, Sidney Brenner, che avrebbe in seguito isolato l'RNA messaggero, Valina. Altri membri

erano Alex Rich, che avrebbe in seguito risolto la struttura tridimensionale del tRNA e scoperto i poliribosomi, e Orgel, che in seguito si sarebbe occupato di origine della vita e del mondo a RNA.

I membri del Club si incontravano 2 volte all'anno e nel frattempo si scrivevano per discutere e comunicare dati che non erano ancora maturi per essere pubblicati su una rivista scientifica. Nel 1955 Crick propose in una sua relazione (fig. 2F2.2) l'**ipotesi dell'adattatore** secondo cui una struttura non ancora scoperta avrebbe portato gli amminoacidi e li avrebbe messi in ordine su una catena di acido nucleico. In questo scritto, che Crick definì successivamente il suo contributo più influente non pubblicato, l'autore ipotizzava anche come poteva essere organizzato il codice e che doveva essere di 64 triplette. Gamow, invece, usava la matematica per stabilire il numero di nucleotidi necessari per fare un codice per 20 amminoacidi, e anche lui giunse alla conclusione che il codice doveva essere fatto di tre lettere.



Figura 2F2.1 Alcuni membri dell'RNA Tie Club. In prima fila A. Rich e J. Watson, in seconda fila F. Crick e L. Orgel.

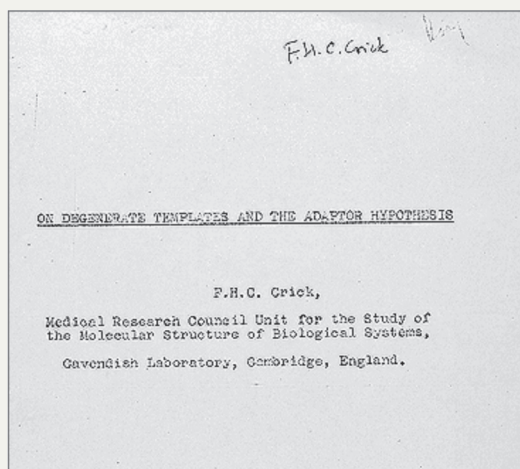


Figura 2F2.2 La nota di F. Crick sull'ipotesi dell'adattatore del 1955.

L'RNA da un punto di vista chimico è molto simile al DNA, con la differenza fondamentale che lo zucchero che costituisce la sua impalcatura è il ribosio, che in posizione 2' ha un gruppo ossidrilico (OH). La presenza di questo gruppo, proprio per l'ingombro sterico che crea, impedisce all'RNA di assumere una conformazione stabile a doppia elica di tipo B. L'RNA è generalmente una catena a singolo filamento, ma è in grado di formare tratti a doppia elica con una conformazione simile al DNA A. L'altra differenza è che contiene l'uracile al posto della timina (fig. 2.43).

Come il DNA, l'RNA ha un ruolo fondamentale nel trasmettere l'informazione genetica, come vedremo quando parleremo di trascrizione e sintesi delle proteine. Inoltre il genoma di molti virus, per esempio l'HIV e l'epatite C (HCV), è costituito interamente da RNA. Anche il genoma di molti virus che infettano le piante ("viroidi") è costituito da forme particolari di RNA a doppia elica.

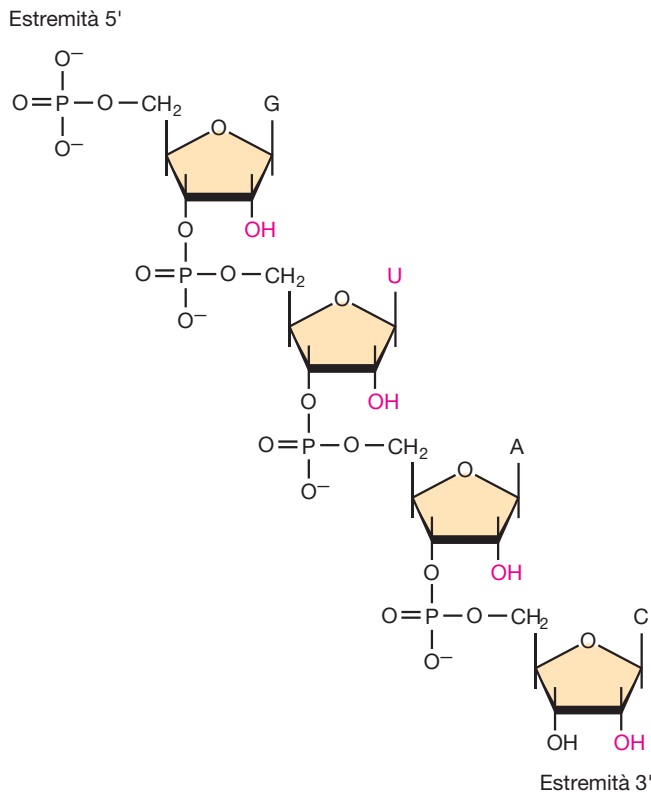


Figura 2.43 Struttura chimica dell'RNA. In rosso sono evidenziate le differenze della struttura chimica dell'RNA rispetto al DNA: (1) presenza di un OH nella posizione 2' dello zucchero; (2) presenza di uracile anziché timina.

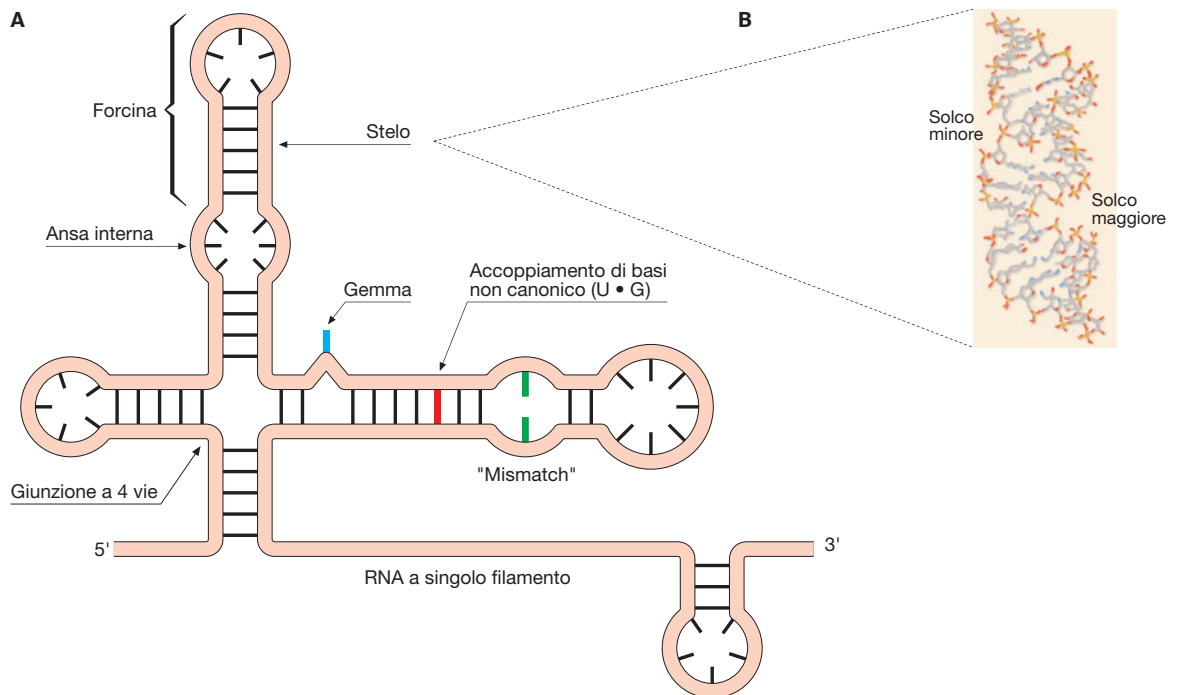


Figura 2.44 Le molecole a singolo filamento dell'RNA possono assumere una varietà di strutture per la presenza di regioni complementari intramolecolari. (A) Sono mostrate alcune delle possibili strutture che prendono il nome di forcine, anse, steli, gemme, anse interne e giunzioni a 4 vie. (B) Struttura tridimensionale di un tratto di RNA a doppia elica dove si può osservare la stretta cavità del solco maggiore e il solco minore più esposto all'esterno, rispetto alla doppia elica del DNA.

gemme, anse interne e giunzioni a 4 vie. (B) Struttura tridimensionale di un tratto di RNA a doppia elica dove si può osservare la stretta cavità del solco maggiore e il solco minore più esposto all'esterno, rispetto alla doppia elica del DNA.

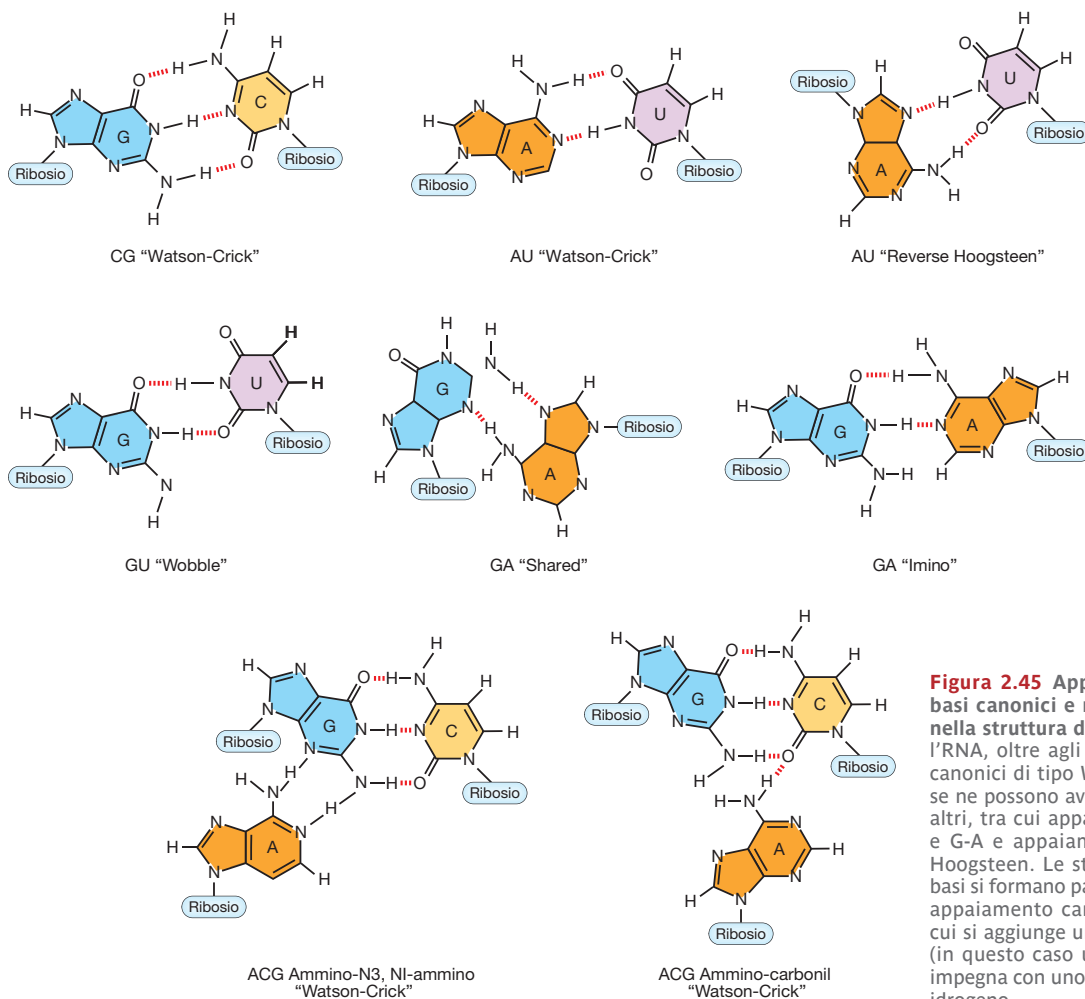


Figura 2.45 Appaiamenti di basi canonici e non canonici nella struttura dell'RNA. Nell'RNA, oltre agli appaiamenti canonici di tipo Watson-Crick, se ne possono avere numerosi altri, tra cui appaiamenti G-U e G-A e appaiamenti di tipo Hoogsteen. Le strutture a tre basi si formano partendo da un appaiamento canonico G-C a cui si aggiunge una terza base (in questo caso una A) che si impegna con uno o due legami idrogeno.

L'RNA ha molteplici funzioni all'interno della cellula e questo è dovuto al fatto che la sua struttura può assumere conformazioni molto diverse che le permettono di diversificare le sue funzioni biologiche. Sequenze diverse consentono la formazione di strutture costituite da anse, forcine, gemme, come illustrato in **figura 2.44**. Questo è possibile per due ragioni: la prima è che le basi possono ruotare e ripiegarsi liberamente intorno al legame fosfodiesterico e ai legami che le legano al ribosio; la seconda è che le basi dell'RNA possono impegnarsi in appaiamenti non soltanto di tipo Watson-Crick, cioè A-U, G-C, ma in un numero diverso di appaiamenti, come per esempio Hoogsteen, incontrati precedentemente nelle strutture di DNA a 3 e 4 eliche (**fig. 2.45**). Gli appaiamenti delle basi nell'RNA sono stati classificati considerando la parte esterna delle basi che partecipano all'interazione e in base all'orientamento dei legami glicosidici relativamente ai legami idrogeno (*cis* o *trans*). I bordi esterni delle basi sono le tre possibili superfici di interazione di un singolo nucleotide, che può essere usato per legami idrogeno di tipo Watson-Crick, Hoogsteen o con gli zuccheri. L'appaiamento Watson-Crick (A-U, G-C) è costituito da legami idrogeno dell'N1 e N6 dell'adenina con l'N3 e O4 dell'uracile e l'N1, N2 e O6 della guanosina con l'N3, O4 e l'N4 della citosina. Tuttavia, nell'adenina il bordo di Hoogsteen implica l'appaiamento tra l'N6 e l'N7. Si possono identificare dodici famiglie di appaiamenti possibili e tutti sono stati descritti nelle strutture cristallografiche di molecole di RNA.

FINESTRA 2.3

STRUTTURA DELL'RNA ED ENERGIA LIBERA

Come abbiamo discusso nel testo (par. 2.4 e fig. 2.44) gli acidi nucleici e in particolare l'RNA, quando sono in forma di singolo filamento, possono generare strutture secondarie formando legami idrogeno tra le basi complementari. In genere le strutture che si formano sono tanto più stabili quanto più lunghi sono i tratti di sequenza complementari appaiati. Data una sequenza nucleotidica relativamente lunga, questa può in teoria essere ripiegata in più possibili strutture secondarie alternative. Quale di queste strutture sia favorita rispetto alle altre, e quindi effettivamente assunta dalla molecola, può essere predetto sulla base del calcolo dell'energia libera contenuta in questa e nelle altre strutture.

La variazione di energia libera, ΔG , permette di prevedere la spontaneità di una reazione chimica (o di una trasformazione di stato) a una certa temperatura e pressione. In altri termini, ΔG è una costante termodinamica che definisce la quantità di energia richiesta o rilasciata da una reazione.

Una reazione che rilascia energia (**esoergonica**) avviene spontaneamente e ha $\Delta G < 0$ (cioè negativo), mentre una reazione che richiede energia (**endoergonica**) ha $\Delta G > 0$ (cioè positivo). Se $\Delta G = 0$, la reazione è all'equilibrio termodinamico. Da un punto di vista termodinamico ΔG è espressa in kcal/mole in base alla sua relazione con la costante di equilibrio K_{eq} :

$$\Delta G = -RT \ln K_{eq}$$

dove R è la costante universale dei gas, T è la temperatura assoluta, $\ln$ è il logaritmo in base e , K_{eq} è la costante di equilibrio. Per la formazione di legami covalenti tra atomi liberi come l'idrogeno e l'ossigeno, i valori di ΔG sono di segno negativo e molto alti, intorno alle -100 kcal/mole, mentre per i legami deboli come quelli di van der Waals e i legami idrogeno sono dell'ordine di -1 , -2 kcal/mole.

Tornando all'applicazione del concetto di ΔG alle strutture formate da molecole di acidi nucleici, l'energia libera globale di una semplice struttura a doppia elica è data dalla seguente equazione:

$$\Delta G_{\text{totale}} = \Delta G_i + \sum \Delta G_x + \sum \Delta G_u$$

dove ΔG_i è l'energia necessaria a formare il primo appaiamento tra due basi per iniziare la formazione di una doppia elica, e avrà sempre un valore positivo; $\sum \Delta G_x$ è la sommatoria di tutti gli appaiamenti successivi che si propagano nella molecola per la formazione della doppia elica e che avranno valori negativi; $\sum \Delta G_u$ è la sommatoria di tutte le strutture non appaiate che si formano perché in quelle regioni le basi non sono complementari e danno origine ad anse, gemme e forcine. Queste hanno in genere valori positivi.

In genere le molecole di RNA a singolo filamento assumono strutture complesse con regioni appaiate a formare doppie eliche che conferiscono stabilità alla struttura, ma anche regioni non appaiate (anse, gemme ecc.) che la destabilizzano.

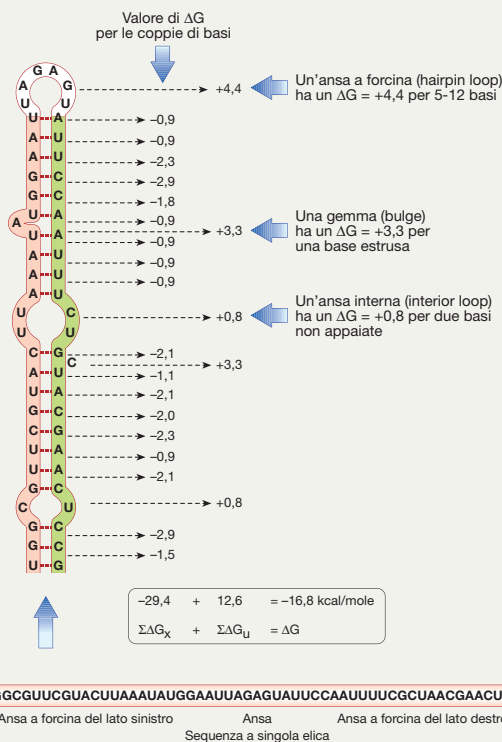


Figura 2F3.1 Calcolo del valore di energia libera di una molecola di RNA strutturata. Una molecola di RNA con la sequenza illustrata in basso, può assumere una conformazione a doppia elica, dove i vari appaiamenti delle basi hanno diversi valori di energia libera. (Fonte: ripresa e rielaborata da B. Lewin, *Genes VI*.)

La stabilità complessiva di una data struttura è descritta dal suo ΔG totale: una struttura con un ΔG negativo più elevato di un'altra è termodinamicamente più stabile, mentre una struttura con un ΔG positivo sarà fortemente instabile e le sue probabilità di formarsi piuttosto scarse.

Prendiamo ora in considerazione un esempio specifico: una molecola di RNA a singolo filamento (**fig. 2F3.1**) ripiegata su se stessa a formare tratti di doppia elica ma contenente anche basi non appaiate (un'ansa terminale, due anse interne e due gemme).

Come si calcola, in pratica, la stabilità di questa particolare struttura? Come mostrato nell'esempio riportato nella figura, il ΔG dell'intera struttura secondaria è uguale alla sommatoria dei singoli ΔG : quelli negativi degli appaiamenti di basi e quelli positivi delle regioni non appaiate (anse, gemme ecc.). Per il calcolo ci si avvale di apposite tabelle che riportano i dati di ΔG , ottenuti sperimentalmente, per i vari tipi di appaiamenti di basi (**tab. 2F3.1**) e altre tabelle, qui non mostrate, che danno i valori di ΔG (positivi) per anse terminali o interne di varie lunghezze, gemme ecc. Si noti che in tabella non viene dato semplicemente il ΔG per

(continua)

Tabella 2F3.1 Tabella dei valori di energia libera associati alle diverse coppie di paia di basi.

Coppie di paia di basi impilate	ΔG
5' – AA – 3' 3' – UU – 5'	–0,9
5' – AU – 3' 3' – UA – 5'	–0,9
5' – UA – 3' 3' – AU – 5'	–1,1
5' – CA – 3' 3' – GU – 5'	–1,8
5' – CU – 3' 3' – GA – 5'	–1,7
5' – GA – 3' 3' – CU – 5'	–2,3
5' – GU – 3' 3' – CA – 5'	–2,1
5' – CG – 3' 3' – GC – 5'	–2,0
5' – GC – 3' 3' – CG – 5'	–3,4
5' – GG – 3' 3' – CC – 5'	–2,9

ciascun appaiamento di basi. Infatti, come abbiamo descritto nel testo parlando di energie di impilamento (par. 2.2 e tab. 2.3), alla stabilità della struttura contribuiscono non solo i legami idrogeno tra le basi appaiate, ma anche le interazioni idrofobiche (impilamento delle basi) che dipendono dalle particolari coppie di basi che si susseguono nella sequenza. Anche queste devono essere tenute presenti quando si creano algoritmi di predizione della struttura secondaria dell'RNA. È per questo che la **tabella 2F3.1** non considera le singole paia di basi, ma coppie di paia di basi, il cui ΔG è in parte dovuto ai legami idrogeno tra le basi appaiate e in parte all'impilamento delle due paia di basi adiacenti.

Queste regole per il calcolo del ΔG sono state proposte da un chimico fisico dell'Università della California a Berkeley, Ignazio Tinoco, che ora sta utilizzando sistemi di nanotecnologie per misurare l'energia contenuta in una singola molecola di RNA. Questi studi si basano sulla possibilità di legare covalentemente le estremità libere 5' e 3' di una molecola di RNA a delle microsferi magnetiche, che vengono poi, con un sistema nanotecnologico, tirate in direzioni opposte. In base alla forza applicata per distendere la molecola, fino a farla diventare un singolo filamento senza strutture, è possibile calcolare l'energia libera reale contenuta in una molecola di RNA con quella particolare sequenza di basi.

Inoltre, come è illustrato in **figura 2.2**, le basi dell'RNA possono essere modificate in molte posizioni e queste modificazioni aumentano le possibilità di impegnare legami idrogeno tra basi in maniera non canonica.

Come abbiamo detto, la presenza dell'ossidrile in posizione 2' impedisce la formazione di una doppia elica di tipo B e favorisce la formazione di un'elica di tipo A dove il solco minore è più largo e accessibile e il solco maggiore più stretto e profondo (**fig. 2.44**). Questa situazione favorisce interazioni di legami idrogeno con il solco minore e rende impossibile la formazione di complessi stabili con proteine nel solco maggiore, come avviene per il DNA. La presenza dell'OH in 2', inoltre, rende la molecola dell'RNA molto reattiva perché il gruppo ossidrili-co tende a fare un attacco nucleofilo sullo scheletro zucchero-fosfato degli acidi nucleici. Questa caratteristica permette all'RNA di avere funzioni catalitiche, come si vedrà nel capitolo sul processamento dell'RNA.

Usando algoritmi che si basano sulle caratteristiche termodinamiche dell'RNA in soluzione e sui vincoli che sono legati agli appaiamenti tra le basi di tipo Watson-Crick o appaiamenti non canonici e il contributo delle basi modificate, le strutture secondarie dell'RNA possono essere predette con una buona approssimazione. Oggi esistono inoltre strutture tridimensionali ottenute con la **cristallografia a raggi X**. La stabilità di una particolare struttura secondaria è funzione di diversi vincoli: il numero di coppie di basi GC verso AU e GU (più è alta l'energia di legame, più stabili sono le strutture), il numero di coppie di basi in uno stelo (più è lungo lo stelo, più legami sono presenti), il numero di coppie di basi in un'ansa a forcina (la formazione di un'ansa con più di 10 o meno di 5 coppie di basi richiede più energia) e infine il numero di basi non appaiate sia come anse interne a un segmento appaiato sia come protuberanze o mismatch (le basi non appaiate diminuiscono la stabilità della struttura).

La stabilità di una struttura secondaria può essere quantificata in base alla quantità di energia libera rilasciata o usata per formare gli accoppiamenti tra le basi.

Un'energia libera positiva richiede del lavoro per assumere una certa configurazione, mentre un'energia libera negativa rilascia lavoro immagazzinato nel sistema. Poiché le energie libere si possono sommare, è quindi possibile calcolare l'energia libera totale di una struttura secondaria, aggiungendo tra loro tutte le componenti di energia libera positive e negative. Tanto più l'energia libera di una struttura sarà negativa, tanto più probabile sarà la formazione di quella struttura (**finestra 2.3**). Gli algoritmi utilizzati calcolano l'appaiamento tra le basi con il minor contenuto di energia. Mentre in genere le anse destabilizzano la struttura, alcune possono essere molto stabili grazie alle interazioni idrofobiche di impilamento che si determinano tra le basi e che hanno un contenuto energetico molto vantaggioso, come per esempio il **tetraloop** formato dalla sequenza GC-UGUU-GU (**fig. 2.46**). Gli appaiamenti tra le basi possono essere formati anche in posizioni non contigue, dando vita a strutture complesse definite **pseudonodi**. Queste strutture si formano quando un segmento di RNA che già ha formato un'ansa in una regione si ripiega e i nucleotidi contenuti nell'ansa interagiscono con i nucleotidi di una regione vicina del filamento. Si viene quindi a formare una seconda regione a doppia elica che dopo una torsione (da qui il termine "pseudonodo") (**fig. 2.47A**) si impila sull'altra. Questo tratto a doppia elica può impegnarsi in legami idrogeno con altre regioni della molecola formando dei tratti a tripla elica (**fig. 2.47B**). Un esempio molto ben studiato di pseudonodo è quello della telomerasi umana, dove questo enzima, che è fondamentale per replicare le regioni telomeriche dei cromosomi, contiene uno stampo di RNA strutturato come uno pseudonodo. La presenza in un complesso enzimatico di un componente di RNA è un concetto fondamentale sul quale torneremo quando, per esempio, parleremo della struttura dei ribosomi. Telomerasi di altri organismi e con sequenze di RNA diverse hanno tutte in comune il motivo a pseudonodo con il quale si ripiega l'RNA, suggerendo un ruolo importante per questa struttura.

Un altro tipo di ripiegamento e interazione intramolecolare dell'RNA è il **motivo A-minore** che si trova per esempio nei ribozimi di gruppo I e negli RNA

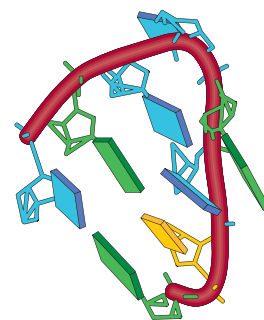


Figura 2.46 Ansa a "tetraloop". Delle varie possibili anse a "tetraloop" qui è mostrata quella formata dalla sequenza GC-UGUU-GU. Per rendere evidente la disposizione spaziale, le basi sono schematizzate con quadrati di diverso colore: G verde, U celeste, C giallo. Sono visibili l'appaiamento non canonico G-U e l'appaiamento canonico C-G, e le quattro basi non appaiate UGUU. La struttura è stabilizzata dalle interazioni idrofobiche di impilamento. Si può anche osservare che le basi al vertice dell'ansa sono esposte all'esterno. Immagine ripresa da RCSB PDB (www.pdb.org) del PDB ID 20J7. [Girard, F.C., Ottink, O.M., Ampt, K.A., Tessari, M., Wijmenga, S.S. (2007), *Nucleic Acids Res*, 35, pp. 2800-2811.]

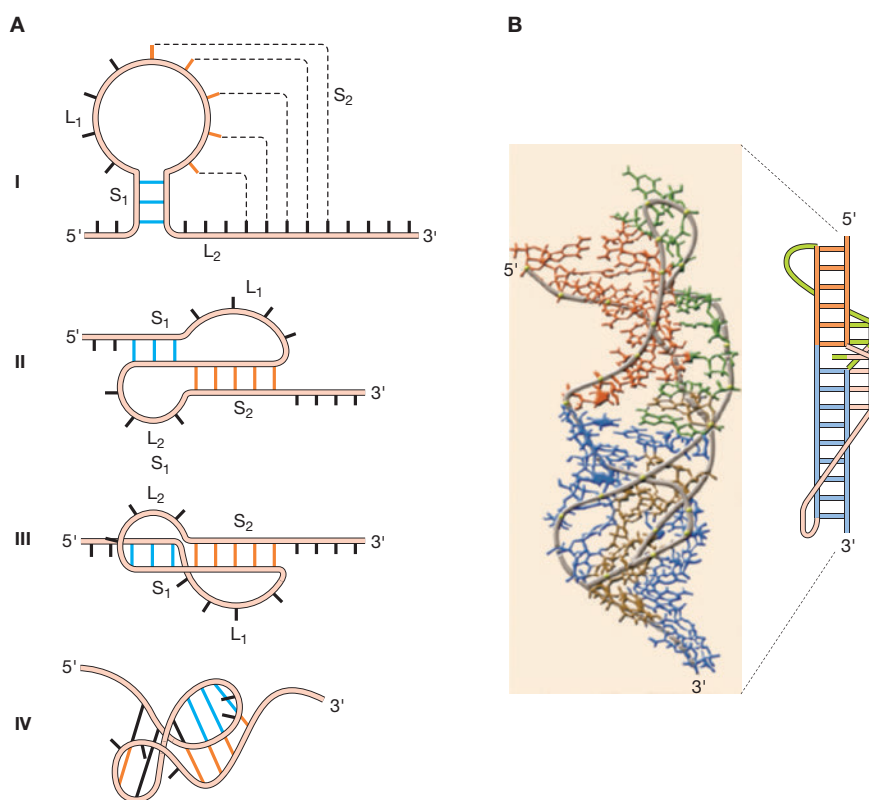
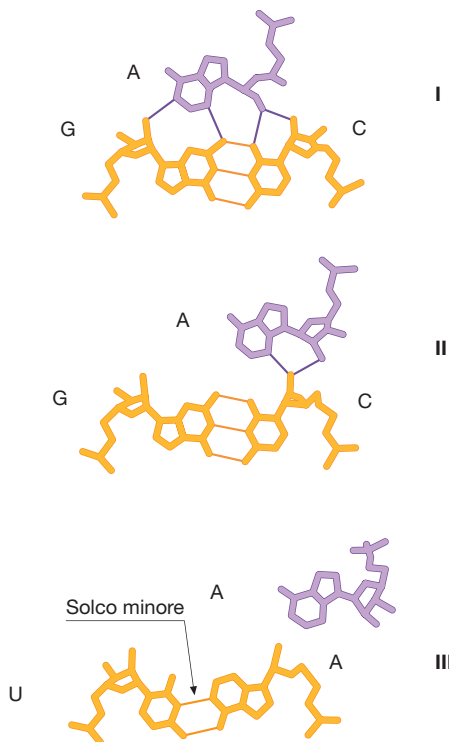


Figura 2.47 Pseudonodo. (A) Formazione di uno pseudonodo: l'ansa di una struttura a forcina (I) interagisce con una regione contigua e forma un secondo tratto a doppia elica (II). La struttura ruota e i due tratti a doppia elica si trovano allineati sullo stesso asse (III e IV). (B) Struttura tridimensionale, generata mediante cristallografia a raggi X, della componente a RNA della telomerasi umana: si può osservare la struttura a pseudonodo e, nella parte centrale, un breve tratto di tripla elica.

Figura 2.48 Esempio di motivo A-minore, una delle strutture più frequenti che si trovano nelle varie specie di RNA. Una adenosina può interagire con il solco minore di una struttura di RNA a doppia elica mediante legami di van der Waals. Qui si vede che l'adenosina si può legare in varie maniere all'interno del solco minore, che nell'RNA a doppia elica è molto accessibile. (I) L'adenina è completamente inserita nel solco minore con cui interagisce molto stabilmente; (II) l'adenina è un po' più esterna al solco minore col quale stabilisce interazioni parziali con lo zucchero e la base; (III) l'adenina è ancora più distanziata dal solco minore con cui interagisce solo tramite un atomo dello zucchero.



ribosomali. Delle adenosine della catena dell'RNA possono interagire specificamente con il solco minore di strutture di RNA a doppia elica mediante legami di van der Waals. In **figura 2.48** si possono osservare i tre diversi tipi di interazioni dell'adenosina con il solco minore e la corrispondente struttura tridimensionale. Il motivo A-minore stabilizza i contatti tra eliche di RNA, interazioni tra anse ed eliche e la conformazione di giunzioni e ripiegamenti molto stretti. L'interazione dell'adenina che si trova sempre al terminale 3' del tRNA con l'RNA ribosomale 23S avviene grazie a un motivo A-minore. Questo tipo di struttura terziaria è la più abbondante nelle interazioni che si formano nella subunità grande del ribosoma.

La risoluzione mediante cristallografia a raggi X di RNA strutturati, come gli RNA ribosomali, di molti ribozimi, di piccoli RNA che costituiscono lo spliceosoma e più recentemente di molti RNA che hanno una funzione regolatrice come i miRNA o i siRNA, ha permesso di definire e catalogare molti tipi di ripiegamenti caratteristici dell'RNA. Una di queste strutture è, per esempio, la *ribose-zipper domain* (chiusura lampo di ribosio) che si forma quando, in un ripiegamento di una lunga regione a doppia elica, due solchi minori si trovano in contatto e il 2'OH di un'elica interagisce con l'ossigeno 2 di una base pirimidinica o con l'azoto 3 di una purina (**fig. 2.49**, dove è rappresentato il dominio P4-P6 dell'introne di gruppo I di *Tetrahymena thermophila*).

Guardando il modo in cui si ripiega l'RNA e le molteplici conformazioni che è in grado di assumere, si può paragonare questa molecola alle proteine. Anche nelle proteine, infatti, i gruppi funzionali si trovano all'esterno e permettono la formazione di un grande numero di strutture diverse che hanno una funzione catalitica. Gli RNA che hanno una funzione catalitica si definiscono ribozimi (se ne parlerà più estensivamente nel paragrafo 12.3 sul processamento dell'RNA).

Proprio per questa capacità dell'RNA di formare strutture diverse, è possibile disegnare catene con sequenza combinatoria e lunghezza variabile. Con questo sistema, da una libreria costituita da miliardi di molecole sintetiche, usando una

Motivo a chiusura lampo di ribosi

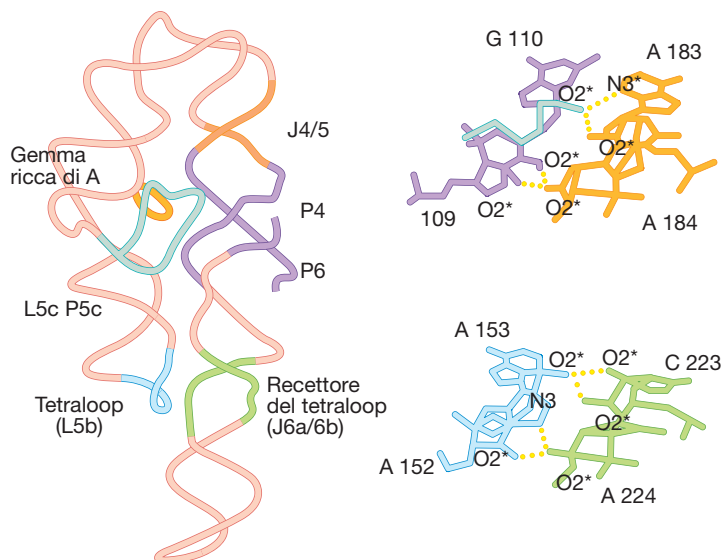


Figura 2.49 Esempio di motivo a chiusura lampo di ribosi: il dominio P4-P6 dell'introne di gruppo I di *Tetrahymena thermophila*. La struttura tridimensionale di questo tratto di RNA mostra come questo sia ripiegato a formare due doppie eliche che interagiscono attraverso i ribosomi, come illustrato a destra.

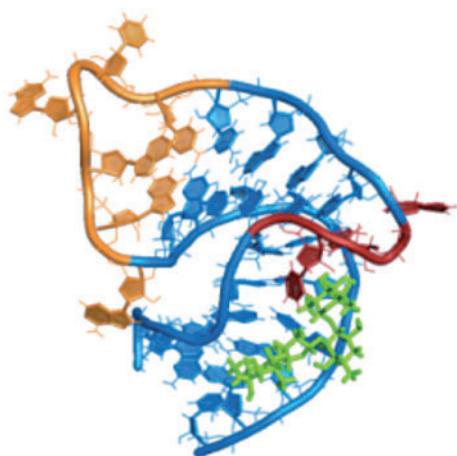


Figura 2.50 RNA aptamero. È mostrato un aptamero formato da due forcine, legato a una molecola di antibiotico (verde).

selezione specifica per la funzione desiderata, si sono generate molecole di RNA che sono in grado di adattarsi e interagire specificamente con un grande numero di molecole diverse, per esempio l'ATP, gli antibiotici (**fig. 2.50**), le proteine o piccoli peptidi e strutture cellulari. Queste molecole di RNA si chiamano **aptameri** perché sono in grado di adattarsi a molecole molto diverse tra loro.

La prima struttura tridimensionale di un RNA a essere risolta con la cristallografia fu quella del tRNA. La sua **struttura a trifoglio** (**fig. 2.51**) fu una vera sorpresa. Ancora più sorprendente fu la risoluzione, nel 1973, della struttura tridimensionale da parte del gruppo di Alexander Rich in cui si vedeva che la struttura a trifoglio ha dei ripiegamenti ulteriori, appaiando tra loro anse distanti e formando una **struttura a L**. La funzione biologica dei bracci delle anse e della struttura generale del tRNA verrà discussa in dettaglio nel paragrafo 14.2 sulla traduzione; qui vengono presentate in breve le caratteristiche generali della struttura. Tutti i tRNA maturi sono mediamente lunghi 73-92 nucleotidi e si ripiegano tutti nella forma a L. I tRNA contengono molte basi modificate. I

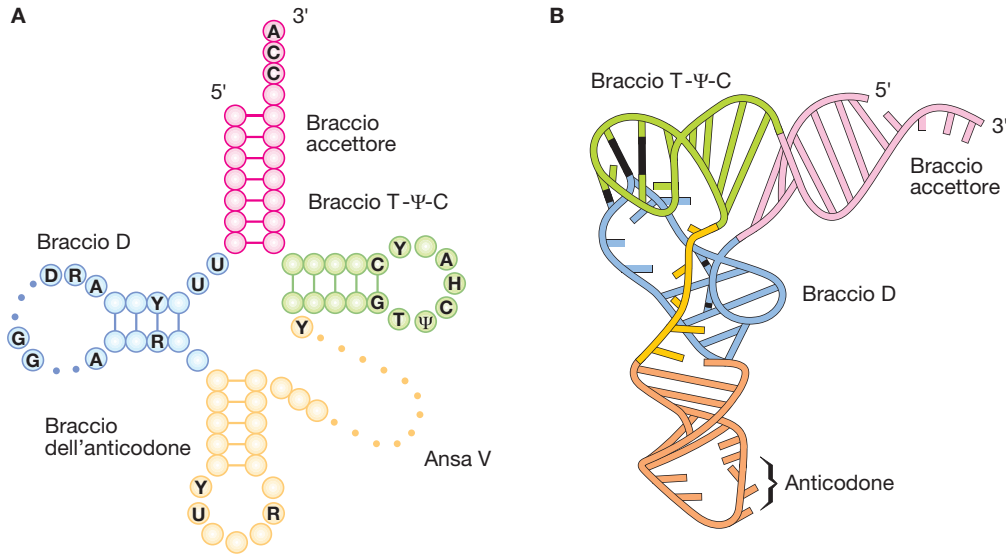


Figura 2.51 Struttura tridimensionale del tRNA. La struttura secondaria a trifoglio (A) si ripiega per appaiamenti di basi tra anse distanti con formazione di una struttura a L (B). I colori aiutano a riconoscere nella struttura a L i domini presenti nella struttura a trifoglio. Si noti come le basi dell'anticodone siano protese all'e-

sterno dell'ansa. Nella figura sono indicati solo i codici delle basi conservate in tutti i tRNA (codici di degenerazione: Y, pirimidina; R, purina; H, adenina, citosina o timina; D, diidrouridina) (può esserne presente più di una D).

motivi strutturali comuni a tutti i tRNA sono le anse che contengono le basi modificate, come l'ansa D che contiene la diidrouridina, l'ansa T-Ψ-C che contiene la pseudouridina, la sequenza CCA all'estremità 3' della molecola. Inoltre in tutti i tRNA avviene una rotazione dei nucleotidi della regione dell'anticodone, che in questa maniera sono esposti all'esterno pur mantenendo tra loro contatti dovuti alle interazioni di impilamento. Un'altra caratteristica importante della struttura è data dall'impilamento coassiale dei bracci che si impegnano in interazioni con regioni non direttamente contigue. Il braccio accettore (dove si lega l'amminoacido) che è costituito da 7 coppie di basi si impila con il braccio T-Ψ-C che è lungo 5 coppie di basi per formare una regione a doppia elica di tipo A lunga 12 nucleotidi, che rappresenta uno dei bracci della struttura a L. Le eliche del braccio dell'anticodone e del braccio D interagiscono tra loro in maniera coassiale per formare l'altro braccio della L. Questo tipo di impilamenti tra steli e anse è comune a molte regioni dell'RNA ribosomale e, come abbiamo visto precedentemente, si trova anche negli pseudonodi.

L'RNA si ripiega e assume strutture terziarie in base alle regole dettate dagli accoppiamenti possibili e dai livelli energetici che queste strutture contengono. Tuttavia, non sempre le conformazioni che esistono *in vivo* e che hanno una funzione biologica sono quelle che da un punto di vista energetico sono le più vantaggiose, e può verificarsi che per un dato RNA la struttura che si forma con grande facilità sia un oggetto completamente inutile da un punto di vista funzionale. Per evitare questo problema l'RNA spesso è associato a proteine che ne regolano e ne promuovono il corretto ripiegamento. Queste proteine appartengono alla famiglia delle hnRNP (ribonucleoproteine nucleari eterogenee) e la loro funzione è quella di "guidare" un RNA verso il suo ripiegamento funzionale. La **figura 2.52** mostra come alcune proteine stabilizzino strutture che devono essere mantenute nella conformazione funzionale, mentre altre, legandosi all'RNA non ancora ripiegato, impediscono l'originarsi di strutture non utili alla formazione della molecola funzionale e si staccano solo quando le regioni che devono interagire sono pronte a farlo. Le proteine ribosomali sono fondamentali per il

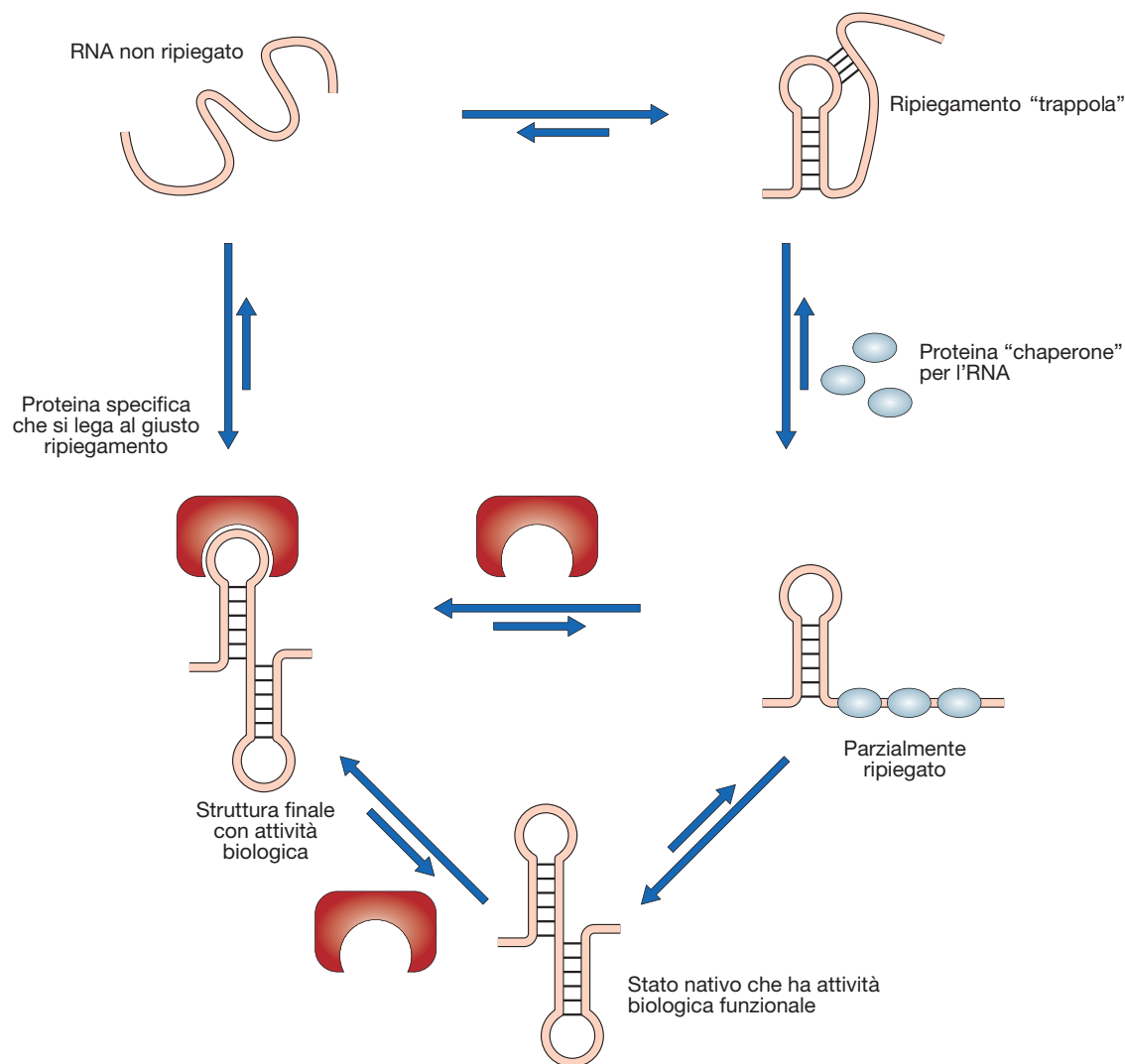
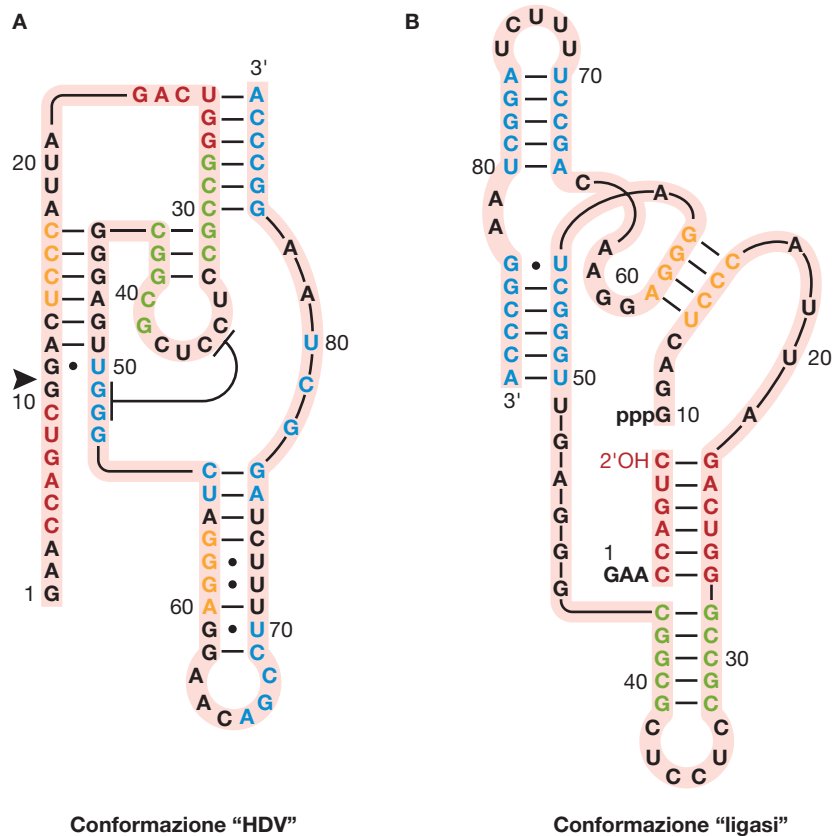


Figura 2.52 Proteine chaperon contribuiscono al corretto ripiegamento delle molecole di RNA. Un RNA potrebbe assumere, anziché la corretta struttura tridimensionale funzionale, una o più strutture alternative, magari favorevoli dal punto di vista energetico, ma non funzionali. A evitare questa eventualità possono intervenire diversi tipi di proteine. Nella figura è mostrata, a sinistra,

una proteina che interagisce con una corretta ansa dell'RNA e la stabilizza impedendone l'interazione con altre regioni dell'RNA. A destra è invece mostrata una proteina che copre una regione a singolo filamento dell'RNA impedendole interazioni improprie, fino a quando essa non interagisce con la sequenza giusta.

ripiegamento degli rRNA, infatti stabiliscono la sequenza temporale degli stadi di ripiegamento dell'rRNA nella formazione delle due subunità del ribosoma e impediscono all'rRNA di formare strutture energeticamente vantaggiose ma biologicamente inutili. Un altro fattore che influenza il ripiegamento è la presenza di ioni che stabilizzano l'RNA, come il Mg^{++} . Nel caso del ribozima a testa di martello, come vedremo nel paragrafo 12.3 sulla maturazione dell'RNA, questo ione bivalente è necessario per la formazione della sua struttura terziaria stabile. Ci sono anche casi in cui da una stessa sequenza di RNA si formano due strutture diverse entrambe funzionali, come nel caso del genoma a RNA del virus dell'epatite delta (**fig. 2.53**). Questo genoma si replica con il meccanismo del cerchio rotante (*vedi* finestra 6.5 e par. 12.3) che genera lunghe molecole di RNA in cui il genoma virale è ripetuto molte volte in tandem. Questo genoma contiene un tratto di RNA di 85 nt che può assumere due strutture alternative con proprietà

Figura 2.53 Un RNA con due strutture alternative con funzioni diverse. Un tratto di circa 90 nt dell'RNA del virus dell'epatite delta può assumere due conformazioni alternative. (A) Conformazione che conferisce all'RNA la funzione di ribozima; è visibile il substrato della reazione e il sito di taglio indicato dalla punta di freccia. (B) Conformazione con attività RNA ligasica di classe III; il sito attivo è composto dalla guanosina trifosfato in posizione 10 e dal 2'OH di una citosina.



funzionali diverse. Una delle due strutture conferisce all'RNA un'attività di ribozima per il taglio dei precursori multimerici in genomi virali singoli (**fig. 2.53A**); la struttura alternativa conferisce all'RNA un'attività ligasica per la circolarizzazione dei singoli genomi virali (**fig. 2.53B**). La cinetica di ripiegamento degli RNA è quindi un fenomeno biologico complesso regolato in maniera accurata.

Lecture di approfondimento

- Dickerson, R.E. (1984), «Come viene letta l'elica del DNA», *le Scienze*, **186**.
- Doudna, J.A. e Cech, T.R. (2002), «The chemical repertoire of natural ribozymes», *Nature*, **418**(6894), pp. 222-228.
- Nelson, J.A. e Uhlenbeck, O.C. (2006), «When to believe what you see», *Molecular Cell*, **23**(4), pp. 447-450.
- Schoeffler, A.J. e Berger, J.M. (2008), «DNA topoisomerases: harnessing and constraining energy to govern chromosome topology», *Quarterly Reviews of Biophysics*, **41**(1), pp. 41-101.
- Wang, J.C. (2002), «Cellular roles of DNA topoisomerases: a molecular perspective», *Nature Reviews Molecular Cell Biology*, **3**, pp. 430-440.
- Wilkins, M.H., Stokes, A.R. e Wilson, H.R. (1953), «Molecular structure of deoxy-pentose nucleic acids», *Nature*, **71**(4356), pp. 738-740.

3

Codice genetico

3.1 Il “dogma centrale” e prime ipotesi sul codice genetico

3.2 Decifrazione del codice genetico

3.3 Fasi di lettura e ORF

3.4 Mutazioni a soppressione e codice genetico

3.5 Struttura del codice genetico e sue proprietà

3.6 Origine ed evoluzione del codice genetico

3.1 Il “dogma centrale” e prime ipotesi sul codice genetico

Una volta dimostrato che il DNA rappresenta il materiale genetico e una volta scoperta la struttura a doppia elica, ci si è posti la domanda di come questa molecola possa svolgere le due funzioni che devono, per necessità, caratterizzare i geni. Infatti, è stato subito chiaro che un gene deve avere due proprietà fondamentali: deve poter essere replicato e deve potersi “esprimere”, cioè determinare o influenzare un carattere fenotipico. Per quanto riguarda la proprietà replicativa del DNA, già nel loro primo brevissimo fondamentale lavoro del 1953 sulla rivista *Nature*, Crick e Watson ipotizzarono un meccanismo di copiatura del materiale genetico che si rivelò corretto (*vedi* par. 6.1 e fig. 6.1). Per quanto riguarda, invece, la proprietà del materiale genetico di esprimere e controllare un fenotipo, il discorso era più elaborato e complesso. Infatti, il modo in cui il patrimonio genetico può dirigere la costruzione di un organismo era, nella prima metà del secolo scorso, un problema avvolto nel mistero più assoluto. Una prima semplificazione del problema è scaturita dall’idea, formulata da Beadle nel 1945, che ciascun gene determina la struttura di una specifica proteina, da cui la nozione che la costruzione di un organismo vivente, per quanto complessa, può essere ridotta alla sintesi regolata delle sue proteine, strutturali ed enzimatiche. Una seconda semplificazione venne successivamente dalla constatazione che, benché la funzione di una proteina sia dovuta alla sua struttura tridimensionale, questa è il risultato del ripiegamento di una sequenza lineare di amminoacidi e che, quindi, la funzione del gene si riduce alla determinazione della sequenza amminoacidica della proteina.

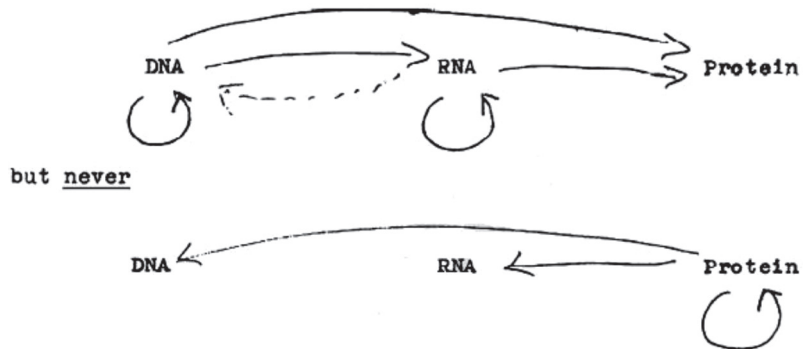
Nel frattempo si era consolidata la nozione di **colinearità gene-proteica**: l’ordine delle mutazioni ottenuto con una mappatura genetica è colineare con l’ordine degli amminoacidi effettivamente cambiati lungo la sequenza amminoacidica della proteina. Questo fa intravedere il gene come una struttura lineare, la sequenza nucleotidica, che codifica per la struttura lineare di una proteina rappresentata dalla sua sequenza amminoacidica: in sostanza, il problema più importante della Biologia poteva ridursi a comprendere un “codice” per passare da una sequenza di 4 diverse basi a una sequenza di 20 diversi amminoacidi.

Figura 3.1 Schema originale di Francis Crick che descrive per la prima volta il "dogma centrale".

The Doctrine of the Triad.

The Central Dogma: "Once information has got into a protein it can't get out again". Information here means the sequence of the amino acid residues, or other sequences related to it.

That is, we may be able to have



where the arrows show the transfer of information.

Il **dogma centrale**, che Crick aveva presentato nel 1958 a un congresso (fig. 3.1), sintetizza le conoscenze di allora sul **flusso dell'informazione genetica**. L'ipotesi, allora non dimostrata e per questo chiamata "dogma", prevedeva che l'informazione genetica può essere trasferita dal DNA all'RNA e dall'RNA alla proteina, oltre che replicata da DNA a DNA e da RNA a RNA, ma che non può essere trasferita dalle proteine di nuovo agli acidi nucleici. Questo concetto ha subito una variazione per quanto riguarda il flusso dell'informazione tra gli acidi nucleici allorché nel 1970 fu scoperta la trascrittasi inversa, un enzima in grado di sintetizzare molecole di DNA da stampi di RNA. Dopo questa notizia Crick scrisse una nota sulla rivista *Nature* per sintetizzare di nuovo i concetti di flusso dell'informazione, ma, ancora oggi, non c'è alcuna evidenza che l'informazione possa passare dalle proteine al DNA. In realtà, come descriveremo nel capitolo 18, oggi sappiamo che alcuni meccanismi epigenetici provocano modificazioni ereditabili dell'espressione genica, senza coinvolgere cambiamenti della sequenza del DNA.

3.2 Decifrazione del codice genetico

A metà degli anni cinquanta si iniziò ad affrontare il problema della decifrazione del codice genetico. Il fatto che questo fosse costituito da triplette di basi era stato presto ipotizzato sulla base di semplici considerazioni numeriche (4 diversi tipi di basi devono codificare 20 tipi diversi di amminoacidi) e di una presunta semplicità del sistema di traduzione che invitava a scartare l'ipotesi di un numero di basi variabile per i diversi amminoacidi. Come abbiamo detto a proposito dell'ipotesi dell'adattatore, Crick, sempre in quella nota per i membri dell'RNA Tie Club, ipotizza che il codice, con grande probabilità, sia costituito da 64 triplette ognuna codificante per un amminoacido. Il codice a triplette fu poi dimostrato sperimentalmente. In un lavoro pubblicato nel 1961, lo stesso Crick, Sidney Brenner e altri, descrivevano "(...) esperimenti di Genetica che, insieme al lavoro di altri, suggeriscono che il codice genetico ha le seguenti caratteristiche generali: a) un gruppo di tre basi codifica per un amminoacido (...); b) il codice non è sovrapposto; c) la sequenza di basi viene letta a partire da un punto d'inizio fisso (...) e non c'è particolare punteggiatura che indichi come selezionare le triplette corrette". Queste

conclusioni erano tratte dai risultati di un'analisi genetica di mutanti del gene *rII* del fago T4, ottenuti mediante trattamento con acridine, sostanze mutagene che inducono delezioni o inserzioni di singoli nucleotidi nella catena polinucleotidica. Mutazioni di questo tipo hanno un effetto molto deleterio perché non provocano il cambiamento di un solo amminoacido nella proteina ma, come mostrato in **figura 3.2B**, mettono il codice fuori fase provocando una lettura scorretta per tutta la porzione della proteina a valle della mutazione stessa. Se poi, mediante ricombinazione, vengono introdotte due di queste mutazioni di tipo opposto (una delezione e una inserzione) abbastanza vicine tra loro, la lettura del codice risulterà sbagliata per il breve tratto compreso tra le due mutazioni ma tornerà normale per la restante parte della proteina codificata dopo la seconda mutazione.

Infine, se mediante ricombinazione viene introdotto un numero crescente di mutazioni dello stesso tipo (o tutte inserzioni o tutte delezioni), si ottiene a un certo punto il ripristino della corretta fase di lettura per la regione proteica a valle del gruppo di mutazioni (**fig. 3.2C**). Questo avverrà quando sarà stato inserito (o rimosso) un numero di basi uguale a quello che codifica per un amminoacido, e questo numero è, appunto, risultato essere tre. La sequenza di tre nucleotidi che codificano per un amminoacido viene chiamata **tripletta** o **codone**; questi due termini sono sostanzialmente sinonimi, anche se “tripletta” fa riferimento alla struttura fisica della molecola, mentre “codone” si riferisce piuttosto all'aspetto informativo.

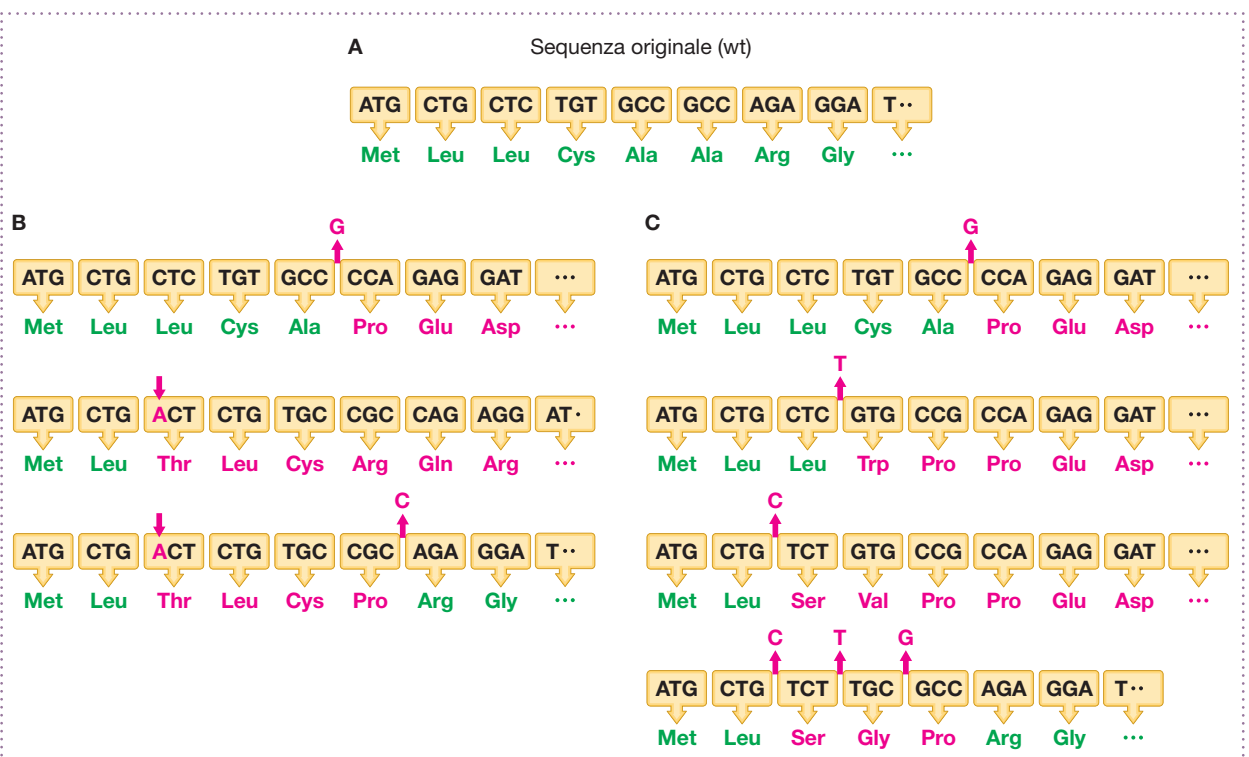
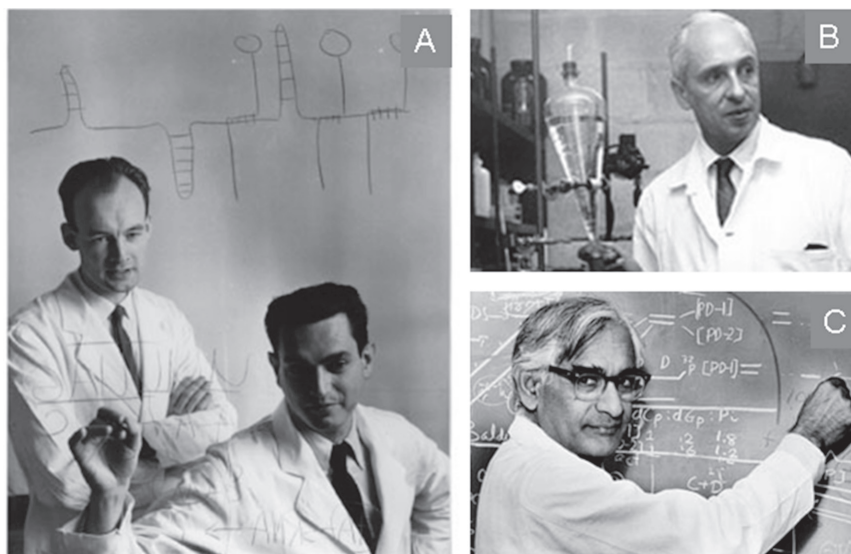


Figura 3.2 Schema semplificato illustrante il principio su cui è basata l'analisi genetica fatta nel 1961 da Crick, Brenner e altri della regione RIIA e RIIB del fago T4 e che dimostrò che il codice genetico è letto a triplette a partire da un punto di inizio fisso e senza “punteggiatura”. (A) Tratto di sequenza di 24 nt codificante per un tratto di 8 amminoacidi di una proteina. (B) La delezione o l'inserzione di singoli nucleotidi (che successivamente sono state chiamate mutazioni “frameshift”) risultano nella sintesi di proteine con sequenza amminoacidica completamente errata a partire dal sito mutato fino alla terminazione. La combinazione di

due mutazioni diverse, una inserzione e una delezione, abbastanza vicine tra loro, risulta in una proteina con amminoacidi errati nel tratto compreso tra i due siti mutati, ma per il resto corretta, in quanto la seconda mutazione ripristina la lettura corretta dei codoni. (C) L'introduzione progressiva di più delezioni (uno, due, tre nucleotidi) come mostrato in questo esempio, o di più inserzioni, ha evidenziato che la lettura corretta viene ristabilita con tre delezioni (o tre inserzioni), dimostrando che il codice genetico è letto a triplette, partendo da un punto di inizio fisso e senza “punteggiatura”.

Figura 3.3 I principali protagonisti della decifrazione del codice genetico negli anni sessanta. (A) Marshall Nirenberg (1927-2010) e Heinrich Matthaei (1929-); (B) Severo Ochoa (1905-1993); (C) Har Gobind Khorana (1922-2011).



La decifrazione del codice genetico, che implicava l'identificazione della corrispondenza tra codoni e amminoacidi, è stata un'importante impresa scientifica che ha visto impegnati vari gruppi di ricercatori, tra cui hanno svolto il ruolo principale Marshall Nirenberg, Heinrich Matthaei, Severo Ochoa e Har Gobind Khorana (**fig. 3.3**). I metodi usati sono stati sostanzialmente due. Il primo metodo, messo a punto da Nirenberg e Matthaei, e mediante il quale nel 1961 venne scoperto che il codone UUU codifica per l'amminoacido fenilalanina, faceva uso di un sistema di sintesi proteica *in vitro* programmata da un mRNA artificiale. Il sistema viene preparato da cellule di *Escherichia coli*, rotte e poi centrifugate per fare sedimentare la frazione particolata contenente frammenti di pareti e di membrane cellulari (**fig. 3.4**). Il supernatante, prelevato dopo la centrifugazione e consistente sostanzialmente del contenuto citoplasmatico della cellula diluito in una soluzione tampone, viene chiamato **estratto cellulare** ed è in grado di portare avanti ancora per un po' di tempo la sintesi proteica che era in corso prima della rottura delle cellule. Questa sintesi però si esaurisce in poco più di dieci minuti, soprattutto per mancanza dell'mRNA, che è il componente più instabile di tutto l'apparato di sintesi proteica. A questo estratto cellulare, ormai privo del suo mRNA endogeno, può essere aggiunto un mRNA artificiale che dirige la sintesi di nuove proteine (**fig. 3.5**). L'efficienza del sistema non è molto elevata e per metterla in evidenza è necessario aggiungere all'estratto amminoacidi marcati radioattivamente (^3H , ^{14}C o ^{35}S), la cui incorporazione nelle proteine è facilmente osservabile e misurabile. È appunto utilizzando questo sistema che Nirenberg e Matthaei scoprirono che il poliribonucleotide sintetico poli(U), costituito da una catena di sole uridine, presenta una certa attività come mRNA e codifica per una catena amminoacidica contenente solo fenilalanina (**fig. 3.6A**). Da qui la conclusione che la tripletta UUU rappresenta il "codone" per tale amminoacido. Il metodo fu poi applicato ad altri omopolimeri, poli(C) e poli(A), e poi a numerosi polimeri misti in cui erano presenti due o anche tre nucleotidi diversi. Con questo metodo fu possibile stabilire solo alcune precise corrispondenze codone/amminoacido, mentre in molti casi si era riusciti a ricavare la composizione dei codoni ma non l'ordine dei tre nucleotidi nei codoni stessi: per esempio, si era potuto stabilire che uno dei codoni contenenti 2A e 1C codifica per treonina, ma non si era potuto identificare quale fosse tra i tre possibili (AAC, ACA, CAA). Infine, la decifrazione del codice fu completata utilizzando un altro metodo che non comporta una vera sintesi proteica *in vitro*. Infatti, era stato osservato che un

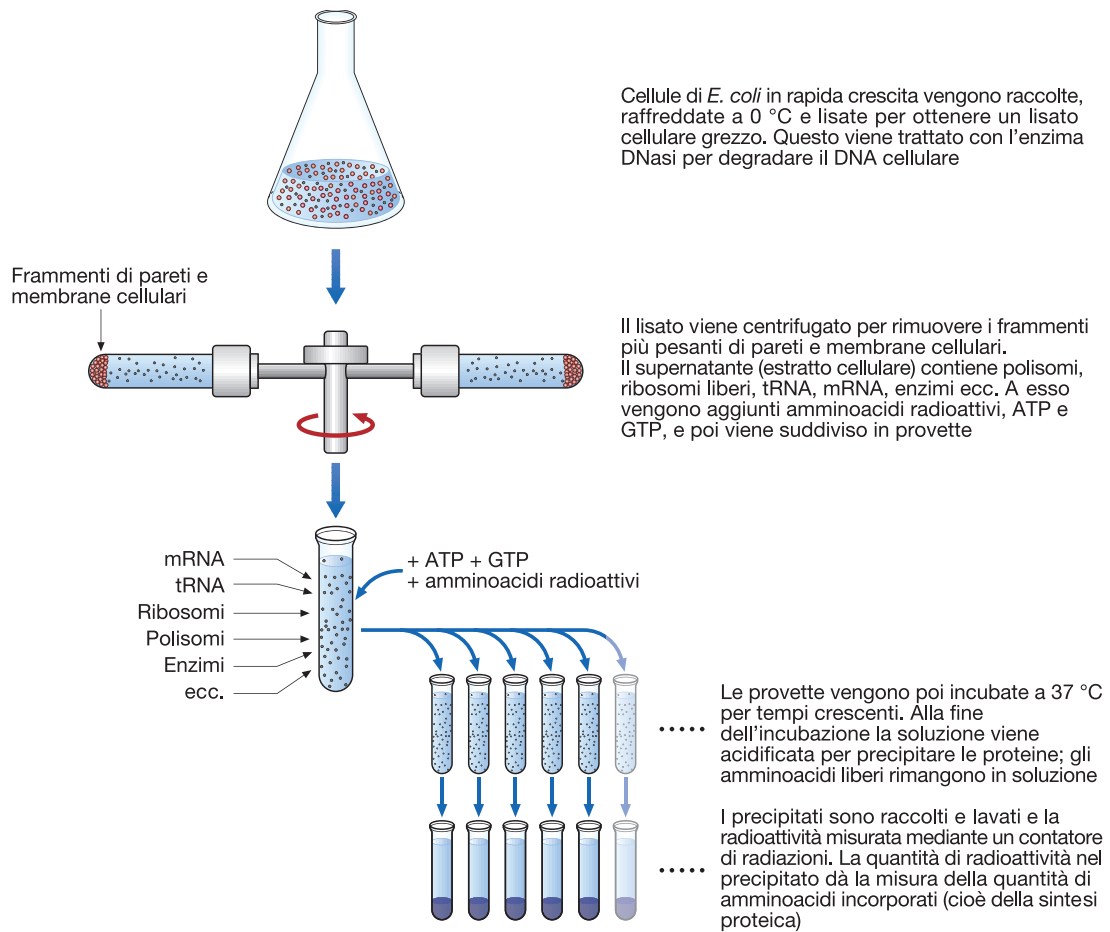


Figura 3.4 Descrizione della preparazione dell'estratto cellulare per la sintesi proteica *in vitro*. Il supernatante, prelevato dopo la centrifugazione, e che consiste del contenuto citoplasmatico del-

la cellula diluito in una soluzione tampone, rappresenta l'estratto cellulare e prosegue ancora per un po' di tempo la sintesi proteica che era in corso prima della rottura delle cellule.

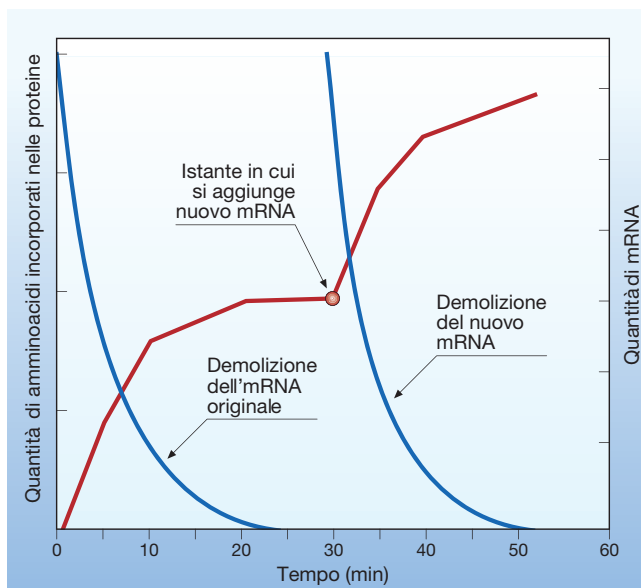
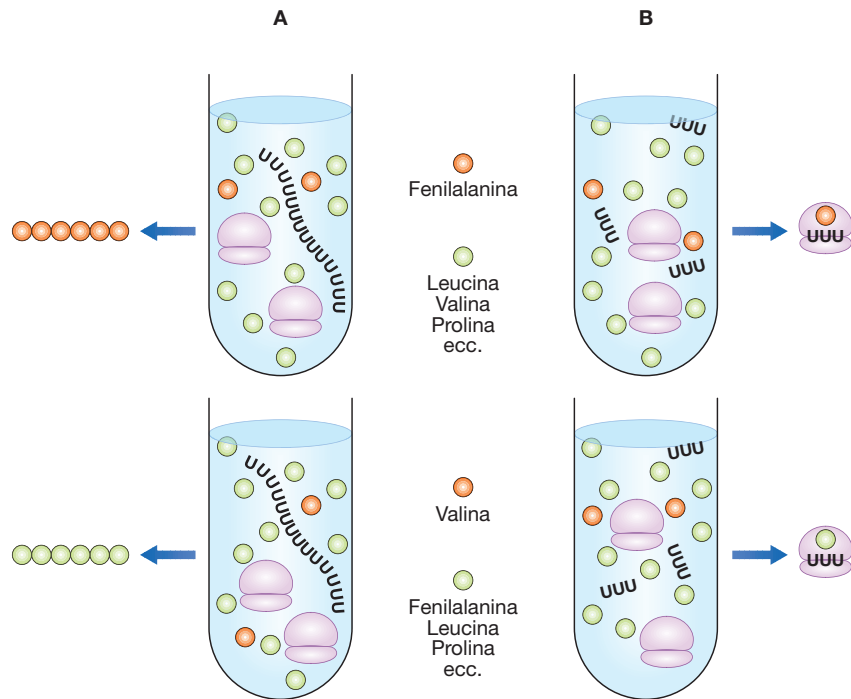


Figura 3.5 Incorporazione di amminoacidi radioattivi nelle proteine sintetizzate in estratto cellulare. La prima parte della curva rossa è dovuta all'mRNA naturale endogeno, la seconda all'mRNA aggiunto.

Figura 3.6 I due metodi che hanno permesso la decifrazione del codice genetico. (A) Il poliribonucleotide sintetico poli(U), costituito da una catena di sole uridine, produce una catena amminoacidica contenente solo fenilalanina. La tripletta UUU, quindi, rappresenta il “codone” per la fenilalanina. (B) Un RNA costituito di una sola tripletta di nucleotidi è in grado di legarsi al ribosoma come un mRNA e di provocare il legame del tRNA corrispondente. Così è stato possibile identificare la corrispondenza codone-amminoacido analizzando quale amminoacido radioattivo veniva legato dal ribosoma in presenza di un dato trinucleotide. Le palline rosse indicano gli amminoacidi marcati radioattivamente, le palline verdi quelli non marcati.



RNA costituito di una sola tripletta di nucleotidi è capace di legarsi al ribosoma come se fosse un mRNA e di provocare il legame del tRNA corrispondente. Per esempio, un trinucleotide ACA si lega al ribosoma e richiama su di esso una molecola di tRNA caricata con la treonina. Così è possibile identificare la corrispondenza codone/amminoacido anche se, essendo l'mRNA costituito da una sola tripletta, non si ha una vera sintesi proteica (**fig. 3.6B**). La scoperta di questo effetto dei trinucleotidi aprì la possibilità di determinare l'ordine dei nucleotidi all'interno dei codoni. Infatti non era difficile a quell'epoca ottenere trinucleotidi sintetici con sequenza desiderata. In questa maniera furono identificati gli amminoacidi codificati da 61 dei 64 possibili codoni costituiti ciascuno da tre basi. I tre rimanenti codoni, che non codificano per nessun amminoacido, vengono chiamati **codoni non senso**; essi segnalano la terminazione della lettura del messaggio e vengono per questo chiamati anche **codoni di terminazione** o **codoni di stop**. La decifrazione del codice genetico fu completata nel 1966, cinque anni dopo l'identificazione del primo codone UUU per le fenilalanina.

3.3 Fasi di lettura e ORF

Gli studi preliminari alla decifrazione del codice portarono anche alla formulazione dei concetti di **fase di lettura** (*reading frame*, tradotto anche come modulo di lettura o cornice di lettura) e di **ORF** (*open reading frame* = fase di lettura aperta). Infatti, come mostrato in **figura 3.7**, la dimostrazione che il codice è costituito di triplette di basi, non è sovrapposto e non è punteggiato, fa sì che ci siano tre modi possibili per tradurre una sequenza nucleotidica in proteina: partendo dal primo nucleotide (fase 1), o dal secondo (fase 2), o dal terzo (fase 3); se partiamo dal quarto nucleotide, dal quinto ecc., ritorneremo sulla prima fase di lettura, poi la seconda ecc. Per definizione, una sequenza nucleotidica che codifica una proteina non contiene codoni di stop in tutta la lunghezza che va dal codone di inizio fino al codone di terminazione; ma ciò è vero solo per quella delle tre fasi di lettura che viene effettivamente decodificata e che chiameremo ORF. Le altre due fasi, che chiameremo

Fase di lettura

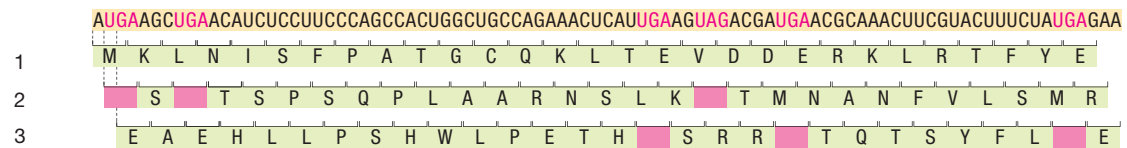


Figura 3.7 Fasi di lettura e ORF. La figura mostra un tratto di 100 nt della parte codificante di un mRNA. Questo può essere letto in tre maniere (fasi): partendo dal 1° nt (fase 1), dal 2° nt (fase 2) o dal 3° nt (fase 3); iniziando dal 4° nucleotide si torna sulla fase 1, e così via. Nella sequenza sono presenti alcuni codoni di terminazione, segnati in rosa. Questi però sono presenti solo in due delle fasi

di lettura (fasi 2 e 3 in questo caso), che quindi non possono codificare per lunghi tratti polipeptidici. La fase di lettura 1, invece, non presenta nessun codone di terminazione e può quindi codificare l'intera proteina terminando all'effettivo codone di stop alla fine della regione codificante. Questa fase di lettura viene chiamata ORF (*open reading frame* = fase di lettura aperta).

A

Gene mutato
contenente un
codone di stop



Trascrizione

mRNA mutato
contenente
un codone di stop



Traduzione

Proteina tronca
non funzionale



B



Gene per un tRNA
con mutazione
nell'anticodone ...

Trascrizione



... che riconosce
il codone di stop

Traduzione



Proteina completa
funzionale

Figura 3.8 Soppressione non senso. (A) Una mutazione che introduce un codone di stop prematuro provoca la formazione di una proteina tronca. (B) Una seconda mutazione nel gene per un tRNA

a livello dell'anticodone può risultare nella sintesi di un tRNA che, riconoscendo il codone di stop, permette l'inserimento di un aminoacido e la prosecuzione della sintesi dell'intera proteina.

chiuse, non avendo questa limitazione possono contenere codoni di terminazione. Poiché, come vedremo, dei 64 codoni, tre sono di terminazione, ce ne aspettiamo, mediamente, uno ogni circa 20 codoni di una sequenza non tradotta. Va notato che le possibili fasi di lettura saranno sei se consideriamo una sequenza di DNA di cui non sappiamo quale dei due filamenti complementari è quello codificante.

3.4 Mutazioni a soppressione e codice genetico

Un ruolo importante nella comprensione del codice genetico lo ha avuto la scoperta dei **mutanti soppressori** o **mutanti Amber** (**finestra 3.1**). Anche questi mutanti sono stati isolati nel sistema del fago T4, selezionando per mutanti condizionali in cui l'effetto di una mutazione molto grave viene soppresso da una seconda mutazione che avviene in un altro gene; nel caso del sistema fago-ospite, la seconda mutazione riguarda un gene dell'ospite. Mutazioni in geni virali facevano diminuire la loro capacità di infezione, generando fagi che erano in grado di infettare e crescere solo in alcuni ceppi di *E. coli*. Quando queste mutazioni sono state caratterizzate, si è scoperto che erano mutazioni che introducevano un codone di stop prematuro provocando la formazione di una proteina tronca (**fig. 3.8**).

FINESTRA 3.1

ORIGINE DEL NOME *AMBER SUPPRESSOR*

Le mutazioni Amber sono state il primo set di mutanti non senso a essere scoperto e sono state isolate nel laboratorio di Richard Epstein al California Institute of Technology alla fine degli anni cinquanta. Il laboratorio di Epstein studiava il fago T4 e, in particolare, la capacità del fago di sopravvivere su ceppi di *E. coli* dopo trattamento con raggi UV. Durante una discussione a cena tra Epstein e il suo collega Charles Steinberg, al quale insistentemente Epstein parlava della sua teoria sui mutanti condizionali (sopravvivere in un ceppo e non in un altro), Steinberg sfidò Epstein a fare il duro lavoro di screening per isolare questi mutanti dopo aver cenato.

Il lavoro consisteva nel prendere diverse migliaia di placche fagiche e piastrarle su due diversi ceppi di *E. coli*. Quella notte convinsero uno studente di dottorato che stava nel laboratorio ad aiutarli nell'esperimento, promettendogli di chiamare il primo mutante eventualmente ottenuto con il suo nome. Lo studente si chiamava Harris Bernstein e in tedesco *Bernstein* vuole dire Ambra (Amber), e quindi il primo mutante che fu isolato dallo screening fu chiamato Amber. I nomi Ochre e Opal sono stati dati successivamente per coerenza con il primo nome assegnato. Più tardi Sidney Brenner identificò i codoni corrispondenti come i tre codoni di stop.

Questa situazione provoca in molti casi una perdita di funzione, poiché non è presente la porzione di proteina a valle della mutazione. Il primo mutante non senso fu chiamato **Amber**. I virus con le mutazioni Amber sono capaci di infettare solo alcuni ceppi di batteri, chiamati appunto **soppressori Amber**. Questi batteri hanno delle mutazioni che permettono il ripristino della funzione virale mutata. Una mutazione in un tRNA a livello dell'anticodone, e che riconosce quindi il codone di stop, permette l'inserimento di un amminoacido in quel sito e ripristina la sintesi completa della proteina. Questi tRNA competono con i fattori di rilascio del ribosoma e permettono quindi l'inserimento di un amminoacido dove prima era presente un codone che per la mutazione era diventato un codone di stop. Quindi i virus che portano una mutazione Amber possono sopravvivere solo in batteri che a loro volta contengono un tRNA soppressore Amber in grado di sopprimere la mutazione nel gene del batteriofago.

Un'altra classe di mutanti che sono stati isolati e che indicavano un arresto della catena è stata quella dei **mutanti Ochre** (ocra), il cui nome è stato dato per analogia ad Ambra. Anche in questo caso questi fagi mutati possono crescere solo in alcuni ceppi "soppressori" di *E. coli*. Con esperimenti di infezione crociata tra i mutanti Amber e Ochre si è visto che questi corrispondevano a una tripletta diversa. Attraverso una serie di esperimenti che comparavano questi mutanti tra loro e con l'analisi dei codoni degli altri amminoacidi, Sidney Brenner è riuscito ad attribuire al mutante Amber la tripletta UAG e al mutante Ochre la tripletta UAA. Il terzo e ultimo codone di stop UGA è stato nominato **Opal**, per coerenza con i primi due.

I mutanti soppressori sui tRNA non sono solo limitati al sistema fago-ospite, ma esistono anche in altri sistemi biologici.

Esistono anche delle mutazioni nei tRNA chiamate **senso**, in cui l'anticodone non corrisponde all'amminoacido caricato. Queste mutazioni sono meno gravi, naturalmente se la sostituzione non comporta una drastica alterazione della conformazione della proteina.

In conclusione possiamo dire che i mutanti soppressori, di non senso e di senso, permettono la sintesi di almeno un po' di prodotto corretto del gene essenziale originalmente mutato. In altre parole, l'introduzione di un certo livello di ambiguità del codice genetico permette la sintesi corretta di parte del prodotto, quanto basta per la sopravvivenza del virus o della cellula. Tuttavia va osservato che, nel contempo, l'ambiguità introdotta dalla mutazione a soppressione risulta in una certa percentuale di errori in tutte le altre proteine sintetizzate. È per questo che i mutanti soppressori, che possono essere isolati e mantenuti in condizioni di laboratorio, non si trovano in natura, essendo svantaggiati nella competizione con ceppi selvatici.

3.5 Struttura del codice genetico e sue proprietà

Il codice genetico è, in genere, rappresentato come mostrato in **figura 3.9**, in modo da metterne in risalto le caratteristiche (si noti che, poiché in realtà il codice viene letto a livello di RNA, troviamo qui usata la base U anziché la T del DNA). Nei riquadri di ciascuna delle quattro righe sono riportati i codoni che iniziano con la stessa base. Nei riquadri di ciascuna delle quattro colonne, invece, sono riportati i codoni che hanno la stessa base in seconda posizione. Infine, ciascun riquadro riporta incolonnati i codoni che differiscono nella terza posizione. È inoltre riportato per ciascun codone o gruppo di codoni l'amminoacido corrispondente.

Il codice genetico decifrato negli anni sessanta e mostrato in **figura 3.9** è quello relativo al batterio *E. coli*, ma ci si è presto resi conto che è esattamente lo stesso in tutti gli organismi, procarioti ed eucarioti, inclusa la specie umana, per cui si dice che *il codice genetico è universale*. Ne consegue che, se trasferiamo un mRNA procariotico in una cellula di mammifero, questo mRNA verrà letto correttamente dall'apparato di traduzione eterologo. In realtà, come vedremo meglio più avanti, a questa "universalità" del codice sono state trovate alcune interessanti eccezioni.

Osservando la figura vediamo che tre dei 64 codoni (UAA, UAG e UGA) non codificano per alcun amminoacido, e sono infatti i cosiddetti **codoni di stop** (o di terminazione o non senso) che segnalano la terminazione della sintesi proteica. Gli altri 61 codoni codificano per i 20 tipi di amminoacidi che vengono utilizzati nella sintesi proteica: ne deriva che un amminoacido può essere codificato da più codoni. Infatti, mentre c'è un solo codone per metionina e triptofano, ce ne sono due per fenilalanina, tirosina, istidina, glutammina, asparagina, lisina, acido aspartico, acido glutammico e cisteina, tre per isoleucina, quattro per valina,

		Seconda posizione				
		U	C	A	G	
Prima posizione (estremità 5')	U	UUU Phe UUC UUA Leu UUG	UCU UCC Ser UCA UCG	UAU Tyr UAC UAA Stop UAG Stop	UGU Cys UGC UGA Stop UGG Trp	U C A G
	C	CUU Leu CUC CUA CUG	CCU CCC Pro CCA CCG	CAU His CAC CAA Gln CAG	CGU Arg CGC CGA CGG	U C A G
	A	AUU Ile AUC AUA AUG Met	ACU Thr ACC ACA ACG	AAU Asn AAC AAA Lys AAG	AGU Ser AGC AGA Arg AGG	U C A G
	G	GUU Val GUC GUA GUG	GCU Ala GCC GCA GCG	GAU Asp GAC GAA Glu GAG	GGU Gly GGC GGA GGG	U C A G

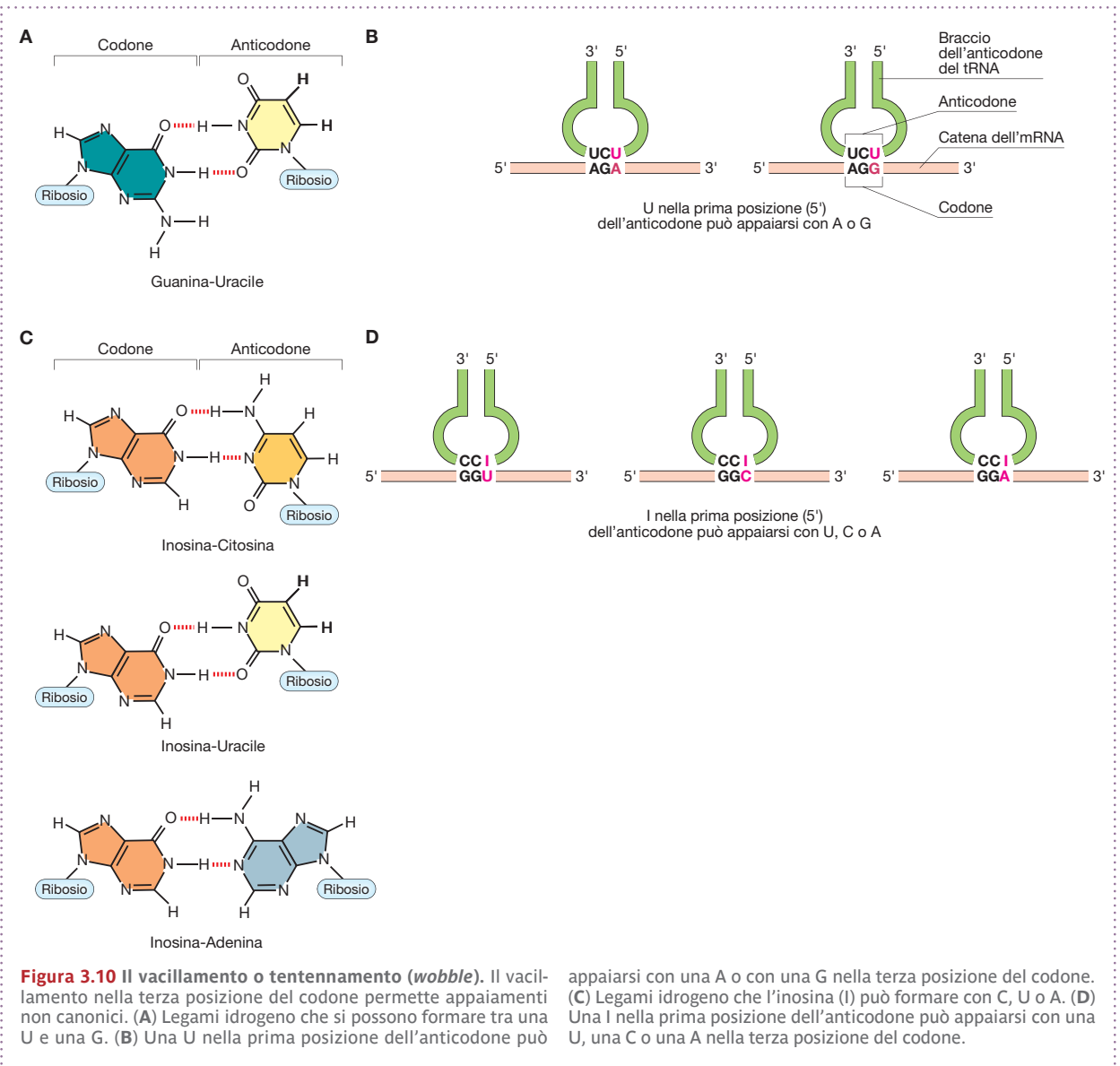
Figura 3.9 Rappresentazione del codice genetico "universale" usato dalla maggior parte degli organismi viventi. Il primo nucleotide di ogni codone è riportato sul lato sinistro della griglia, il secondo in alto e il terzo a destra. Le sequenze dei codoni sono scritte in direzione 5' → 3'. 61 dei 64 codoni codificano per amminoacidi e tre (in rosso) sono codoni di terminazione. È evidente la "degenerazione" del codice, per cui un amminoacido è codificato da più di un codone.

prolina, treonina, alanina e glicina e infine ben sei per leucina, serina e arginina. I diversi codoni che specificano per lo stesso amminoacido vengono chiamati **codoni sinonimi** e un codice così fatto (più parole con uno stesso significato) viene detto **codice degenerato**. La degenerazione del codice genetico, come per tutti i codici, non pone problemi di specificità nella lettura del messaggio.

D'altra parte la figura mostra anche che *il codice genetico non è ambiguo*, cioè ciascun codone codifica per un solo amminoacido, senza eccezioni. Questo è importante perché un'ambiguità del codice, a differenza della degenerazione, non permetterebbe una lettura univoca del messaggio e avrebbe gravi conseguenze sulla specificità della sintesi proteica. In realtà, abbiamo visto che le mutazioni tRNA soppressori introducono un livello di ambiguità e possono essere mantenute in laboratorio, ma saranno comunque svantaggiose in natura.

Osserviamo ora in dettaglio in **figura 3.9** la collocazione dei diversi amminoacidi e dei diversi codoni. Per quanto riguarda gli amminoacidi, si può notare che spesso, anche se non sempre, amminoacidi con proprietà chimiche simili si trovano nella stessa area della tabella, il che vuol dire che sono specificati da codoni abbastanza simili tra loro. Questa osservazione, insieme al fatto che molti codoni che differiscono in terza posizione specificano per lo stesso amminoacido, ci suggerisce l'idea che il codice sia fatto in modo che, in molti casi, le mutazioni di singole basi o non comportano alcun cambiamento a livello di sequenza amminoacidica, oppure portano a cambiamenti conservativi, cioè alla sostituzione di un amminoacido con un altro simile, minimizzando così l'effetto dirompente delle mutazioni. Per quanto riguarda la collocazione dei codoni si nota che in molti casi (8 su 16) i quattro codoni che si ritrovano in un riquadro codificano per lo stesso amminoacido. In altri 6 casi i primi due codoni codificano per un amminoacido e i secondi due per uno diverso. Considerato che, per come è strutturata la tabella, i codoni di ogni riquadro differiscono solo nella terza base del codone, possiamo concludere che, in genere, la terza base dei codoni non è coinvolta nella codificazione dell'amminoacido, o al massimo è coinvolta come pirimidina o purina, senza però distinzione tra C e U o tra A e G. Come approfondiremo in dettaglio altrove, la decodificazione del messaggio genetico viene fatta dai tRNA, che con i loro "anticodoni" riconoscono, per appaiamento di basi complementari, i codoni. In linea di principio ci si sarebbe potuti aspettare che fossero necessari 61 diversi tRNA per riconoscere i 61 codoni, anche se gli amminoacidi da codificare sono solo venti. La degenerazione del codice in terza posizione, però, insieme al principio generale di parsimonia, suggeriva che in genere nel riconoscimento codone-anticodone siano implicati solo due appaiamenti di basi (la prima e la seconda del codone con, rispettivamente, la terza e la seconda dell'anticodone). Non sarebbe quindi coinvolto un appaiamento specifico tra la terza base del codone e la prima dell'anticodone, o al massimo ci potrebbe essere un riconoscimento generico purina-pirimidina. Fu quindi proposta da Crick la **teoria del tentennamento o vacillamento** (in inglese *wobble*), per cui nell'appaiamento codone-anticodone nel sito di decodificazione del ribosoma, l'ansa dell'anticodone del tRNA presenta una certa deformazione strutturale. Così, mentre nelle prime due posizioni del codone vale la solita regola di appaiamento C-G e A-U, nella terza posizione è possibile, oltre a questo, anche un appaiamento anomalo G-U. Questo fenomeno risulta nello schema di appaiamento delle basi mostrato in **figura 3.10** e può spiegare, da una lato, la degenerazione del codice in terza posizione e, dall'altro, la necessità di un numero di tRNA inferiore a 61.

Successivamente è stato trovato che alcuni tRNA hanno, nella prima posizione dell'anticodone, delle basi modificate quali l'inosina e la tiouridina che hanno proprietà di appaiamento di basi particolari e potrebbero così contribuire alla lettura del codice degenerato in terza posizione con un numero ridotto di tRNA. La base strutturale di questo fenomeno è dovuta, oltre alla capacità delle basi modificate di impegnarsi in appaiamenti non canonici, anche alla maggiore



libertà di movimento della prima base dell'anticodone, a causa delle interazioni di impilamento. Infatti, come si può vedere in **figura 3.11**, le basi dell'anticodone sono sottoposte a interazioni di impilamento, anche se protrudono all'esterno del loop dell'anticodone. L'ultima base dell'anticodone si trova all'interno di una pila di 4 nucleotidi, di cui generalmente una è una purina modificata che ha un notevole ingombro e ne limita i movimenti. La prima base, invece, è alla fine della pila ed è più libera di muoversi, permettendo, quindi, di "tentennare". Sono necessari almeno 31 tRNA, oltre al tRNA d'inizio, per decodificare tutti i 61 codoni.

3.6 Origine ed evoluzione del codice genetico

Dopo la decifrazione del codice genetico ci si è domandati se il codice è fatto così per necessità o per caso. Per esempio, il fatto che il codone GCA codifichi per alanina è dovuto a una qualche affinità chimica tra le due strutture che li avrebbe in qualche modo messi in relazione tra loro ai primordi della vita sulla Terra, op-

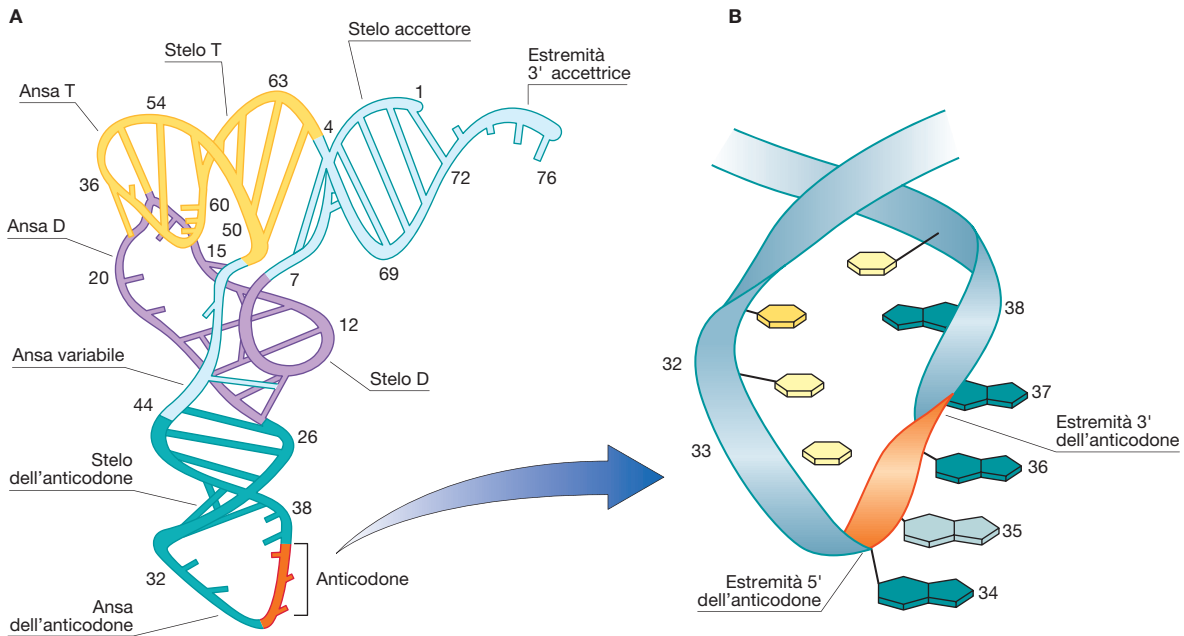


Figura 3.11 Le basi strutturali del vacillamento. (A) La struttura generale del tRNA con le basi dell'anticodone colorate in rosso. (B) L'ansa dell'anticodone in dettaglio. Le basi dell'anticodone sono estruse all'esterno, ma sono comunque impegnate in interazioni di impilamento tra loro. La terza base (36) si trova all'interno di

una pila di 4 basi di cui generalmente una è una purina modificata che ha un notevole ingombro e ne limita, quindi, i movimenti. La prima base dell'anticodone (34), invece, è alla fine della pila ed è più libera di muoversi, permettendo, quindi, di "vacillare".

pure è una relazione nata casualmente che si è però consolidata durante le prime fasi dell'evoluzione? Nonostante i molti esperimenti fatti, non è stata dimostrata in modo del tutto convincente una affinità tra gli amminoacidi e i codoni che li codificano o i nucleotidi che li compongono. Questo fa quindi pensare che il codice genetico attuale sia il risultato di un congelamento di un codice primordiale basato sul caso. In altre parole e per assurdo, se potessimo tornare indietro e fare ri-iniziare, nel brodo primordiale che si era formato sulla Terra quattro miliardi di anni fa, una vita basata di nuovo sul binomio acidi nucleici-proteine con le stesse basi e gli stessi amminoacidi, il codice genetico potrebbe essere diverso da quello che abbiamo oggi. Pur non essendovi prove certe, si pensa che all'origine della vita un ruolo molto importante fosse giocato dall'RNA (vedi par. 18.3). Questo, infatti, è in grado di svolgere il ruolo di genoma analogamente al DNA, ma anche, al tempo stesso, di avere attività catalitiche nel metabolismo dell'RNA stesso (tagli, saldature, modificazioni). In uno scenario di questo genere, il codice inizialmente sarebbe stato abbastanza lasso, ambiguo, ma poi la pressione selettiva lo avrebbe portato a una sempre maggiore specificità e al tempo stesso a un maggiore irrigidimento. A un certo punto, l'aumentata complessità della cellula primordiale con un numero di geni sempre maggiore avrebbe reso il codice non più modificabile. Infatti, un cambiamento del codice avrebbe effetti fortemente pleiotropici, che porterebbero cioè al cambiamento di tutte le proteine codificate da un genoma, e quindi sarebbe sicuramente letale. Solo genomi molto piccoli che codificano per pochissime proteine potevano sopportare, anche se molto raramente, minimi cambiamenti di codice. In conclusione si pensa che il codice si sia formato così com'è per caso ma che poi sia rimasto *congelato* durante gran parte dell'evoluzione dalla cellula primordiale agli eucarioti superiori. Piccole differenze marginali sono potute avvenire in genomi molto piccoli, quali quelli dei mitocondri, che possono adottare diverse varianti del codice genetico universale in differenti raggruppamenti.

menti tassonomici (vertebrati, ascidie, nematodi, protisti ecc.). Per esempio il codice genetico mitocondriale adottato nei vertebrati è rappresentato in **figura 3.12**. È importante notare che questo codice semplificato richiede solo 22 tRNA invece dei 31 necessari per il codice universale. In alcune specie si osservano deviazioni rispetto al codice universale anche nei geni nucleari, anche se queste sono molto meno frequenti di quelle osservate nei mitocondri. Per esempio, in alcune specie di ciliati, diplomonadi o alghe verdi UAG e spesso anche UAA codificano per la glutammina, e UGA codifica per la cisteina. In alcune specie di funghi, come nel genere *Candida*, CUG codifica per la serina.

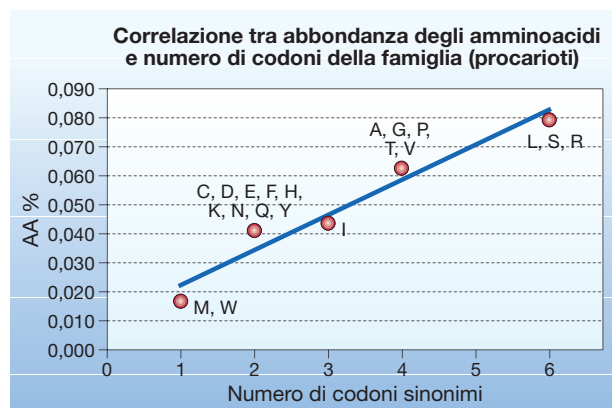
Finora abbiamo descritto il codice per 20 amminoacidi, ma oggi sappiamo che esistono anche altri due amminoacidi: la selenocisteina e la pirrolisina. La selenocisteina è un importante componente di molte proteine coinvolte in reazioni redox. Inoltre, le proteine che contengono la selenocisteina sono fondamentali per lo sviluppo dei mammiferi. La selenocisteina è codificata dal codone di stop UGA. L'ultimo amminoacido, la pirrolisina, è stato identificato recentemente negli archeobatteri ed è associato al codone di stop UAG (*vedi* finestra 14.1).

Il codice genetico si è evoluto per rendere massima l'efficienza della traduzione e per minimizzare gli effetti deleteri delle mutazioni. Il fatto che è degenerato rende i geni meno vulnerabili alle mutazioni casuali. Un codone quattro volte ridondante può subire qualsiasi mutazione alla sua terza posizione e un codone due volte ridondante può subire una delle tre possibili mutazioni alla sua terza posizione senza che l'amminoacido da esso espresso, e di conseguenza la struttura della proteina dove viene inserito, cambi del tutto. Le transizioni (da una purina all'altra o da una pirimidina all'altra) più frequenti in tutte e tre le posizioni del codone producono cambiamenti meno drastici delle trasversioni (da purina a pirimidina o viceversa), che sono meno frequenti.

		Seconda posizione				
		U	C	A	G	
Prima posizione (estremità 5')	U	UUU Phe (GAA) UUC UUA Leu (UAA) UUG	UCU Ser (UGA) UCC UCA UCG	UAU Tyr (GUA) UAC UAA Stop UAG Stop	UGU Cys (GCA) UGC UGA UGG Trp (UCA)	U C A G
	C	CUU Leu (UAG) CUC CUA CUG	CCU Pro (UGG) CCC CCA CCG	CAU His (GUG) CAC CAA Gln (UUG) CAG	CGU Arg (UCG) CGC CGA CGG	U C A G
	A	AUU Ile (GAU) AUC AUA Met (CAU) AUG	ACU Thr (UGU) ACC ACA ACG	AAU Asn (GUU) AAC AAA Lys (UUU) AAG	AGU Ser (GCU) AGC AGA Stop AGG Stop	U C A G
	G	GUU Val (UAC) GUC GUA GUG	GCU Ala (UGC) GCC GCA GCG	GAU Asp (GUC) GAC GAA Glu (UUC) GAG	GGU Gly (UCC) GGC GGA GGG	U C A G

Figura 3.12 Codice genetico mitocondriale nei vertebrati. Vi sono quattro differenze rispetto al codice universale: AUA (Ile > Met) e UGA (Stop > Trp) evidenziate in verde, e AGA (Ser > Stop) e AGG (Ser > Stop) evidenziate in rosso. Tra parentesi è indicato il relativo anticodone. Si noti che ci sono quattro codoni di stop.

Figura 3.13 Evoluzione del codice genetico. Il codice genetico si è evoluto per rendere massima l'efficienza della traduzione e per minimizzare gli effetti delle mutazioni: le transizioni (più frequenti) in tutte e tre le posizioni del codone producono cambiamenti meno drastici delle trasversioni (meno frequenti). Gli amminoacidi sono indicati con la simbologia a lettera singola.



La **figura 3.13** illustra la correlazione, nei procarioti, tra la frequenza dei diversi amminoacidi nelle proteine e il numero dei codoni corrispondenti. Per esempio l'amminoacido molto raro triptofano (W) è rappresentato da un solo codone, mentre la leucina (L), che è l'amminoacido più frequente nelle proteine, è rappresentata da sei codoni. In realtà questa correlazione è ben evidente solo se, come abbiamo fatto nel grafico, consideriamo i valori medi per gruppi di amminoacidi con lo stesso numero di codoni, mentre non è sempre valida per ciascun singolo amminoacido.

Lectures di approfondimento e siti web

- Brenner, S., Stretton, A.O., Kaplan, S. (1965), «Genetic code: the 'nonsense' triplets for chain termination and their suppression», *Nature*, **206**(988), pp. 994-998.
- Crick, F.H. (1966), «Codon-anticodon pairing: the wobble hypothesis», *Journal of Molecular Biology*, **19**(2), pp. 548-555.
- Crick, F.H. (1968), «The origin of the genetic code», *Journal of Molecular Biology*, **38**(3), pp. 367-379.
- Vetsigian, K., Woese, C. e Goldenfeld, N. (2006), «Collective evolution and the genetic code», *Proc. Natl. Acad. Sci. USA*, **103**(28), pp. 10696-10701.
- Il sito «The Genetic Codes in NCBI» (<http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi>) riporta le 24 varianti del codice genetico universale rilevate nelle sequenze collezionate nella banca dati Genbank dell'NCBI.

B

ORGANIZZAZIONE ED EVOLUZIONE DI GENI, CROMOSOMI E GENOMI

Il materiale genetico, generalmente costituito da una o più molecole di DNA, deve essere strutturato e organizzato all'interno della cellula in modo tale da assicurare la massima accuratezza nella trasmissione alle cellule figlie e garantire un adeguato grado di accessibilità per permettere lo svolgimento del corretto piano di espressione genica nel giusto contesto spazio-temporale. Da un lato vi è la necessità di compattare il più possibile i lunghi filamenti di DNA perché possano essere alloggiati all'interno della cellula e trasmessi fedelmente alle cellule figlie, dall'altro quella di garantire l'accessibilità agli apparati replicativo e trascrizionale, che rispondono alle esigenze della propagazione dell'informazione genetica e dell'attuazione del piano di espressione genica in ciascun organismo vivente.

Nel corso dell'evoluzione, diverse soluzioni sono state elaborate per l'impacchettamento dei genomi virali, procariotici ed eucariotici, che saranno sommariamente descritti nel primo capitolo di questa sezione e che comportano il coinvolgimento di diverse proteine. In particolare, descriveremo l'organizzazione e l'impacchettamento del DNA eucariotico, che, essendo di dimensioni molto maggiori di quello virale o procariotico, risulta più problematico. Illustreremo, quindi, la struttura della cromatina e dei suoi componenti elementari, i nucleosomi, e l'implicazione delle sue modificazioni nei processi di regolazione genica.

Utilizzando il linguaggio delle tecnologie dell'informazione possiamo immaginare il materiale genetico come un software molto particolare, in quanto è in grado di guidare la costruzione del proprio hardware, ovvero l'organismo vivente. Questo programma software è anche soggetto a frequenti aggiornamenti che ottimizzano le prestazioni dell'hardware nelle diverse circostanze in cui esso si trova a operare. L'evoluzione, attraverso i processi di mutazione, selezione e fissazione, è responsabile di questi "aggiornamenti".

La possibilità di determinare la sequenza dei nucleotidi nel DNA ci ha permesso di leggere le "istruzioni" presenti nel software. Nel secondo capitolo di questa sezione verranno descritte le diverse caratteristiche dei genomi e dei geni

procariotici ed eucariotici, anche alla luce dell'enorme quantità di dati ottenuti nell'ambito dei progetti di sequenziamento genomico. Vedremo, quindi, come, soprattutto negli eucarioti, solo una piccolissima parte del genoma è utilizzata per fornire le informazioni per la costruzione dei pezzi dell'hardware, ovvero le sequenze codificanti proteine. È andato via via emergendo come la porzione non codificante del genoma, costituita per lo più da sequenze ripetitive, come pure la struttura discontinua dei geni eucariotici, siano fondamentali in molti processi di regolazione dell'espressione genica, sia a livello trascrizionale che post-trascrizionale, e garantiscano un'elevata plasticità e capacità evolutiva dei genomi.

4

Impacchettamento del DNA genomico: cromosomi e cromatina

4.1 Impacchettamento dei genomi virali

4.2 Impacchettamento del genoma batterico

4.3 Organizzazione e impacchettamento del DNA eucariotico

4.4 Cromatina interfase e cromosomi mitotici

4.5 Organizzazioni atipiche di cromosomi e genomi

4.6 Centromeri

4.7 Telomeri

4.8 Nucleosomi e proprietà strutturali e funzionali della cromatina

Il materiale genetico di qualsiasi organismo, procariotico ed eucariotico, è organizzato in strutture superiori, i cromosomi, localizzati spazialmente in una regione specifica della cellula. Un problema che può essere immediatamente riscontrato è la lunghezza delle molecole di DNA in rapporto alle dimensioni del compartimento o struttura cellulare che le contiene. Per esempio, come indicato in **tabella 4.1**, il genoma di *Escherichia coli* è rappresentato da un solo cromosoma costituito da un'unica molecola circolare di DNA comprendente circa 4×10^6 pb, per una lunghezza lineare di circa 1 mm. Tale molecola si trova compattata nel **nucleoide** della cellula batterica, la cui lunghezza è di circa 1 μm , cosa possibile essendo la molecola di DNA molto sottile: ricordiamo che il diametro della doppia elica è di 20 Å, cioè 2 nm.

Ancora più problematica è l'organizzazione dei genomi degli organismi eucariotici. Per esempio, la stessa **tabella 4.1** mostra che il genoma umano è costituito da un corredo diploide di 23 coppie di cromosomi, per un totale di circa 6×10^9 pb. Se le molecole di DNA di una singola cellula fossero distese e allineate, coprirebbero una lunghezza di circa 2 m. Esse si trovano tuttavia compattate nel nucleo cellulare, che ha un diametro di circa 6 μm .

Per esprimere quantitativamente il grado di compattamento delle molecole di DNA si utilizza il valore di **quoziente di compattamento** (o **impacchettamento**), cioè il rapporto tra la lunghezza del DNA e le dimensioni della struttura o compartimento che lo contiene. Per esempio, nel caso di *E. coli* abbiamo un quoziente di impacchettamento di circa 1000 (1000 $\mu\text{m}/1 \mu\text{m}$), mentre in un cromosoma mitotico umano il quoziente di impacchettamento raggiunge il valore di circa 10 000 (si consideri che un genoma diploide umano è costituito da 6×10^9 pb, corrispondenti a circa 2 m di lunghezza totale, suddivisi in 46 cromosomi, ciascuno della lunghezza media di 4,4 cm che deve essere contenuto nel nucleo di circa 6 μm di diametro).

Questa enorme lunghezza delle molecole di DNA rispetto alle strutture cellulari pone un rilevante problema topologico, soprattutto al momento della duplicazione cellulare. Infatti, si può facilmente immaginare come la lunghezza delle molecole possa creare dei problemi al momento della segregazione delle due co-

Tabella 4.1 Caratteristiche del materiale genetico di diversi organismi.

	Struttura contenitore		Genoma			
	Forma	Dimensioni	Acido nucleico*	Lunghezza		
TMV	Filamento	0,30 × 0,02 μm	1 RNA sf	2 μm	6,4	kb
Fago fd	Filamento	0,88 × 0,01 μm	1 DNA sf	2 μm	6	kb
Adenovirus	Icosaedrico	0,09 × 0,09 μm	1 DNA de	12 μm	35	kb
Fago T4	Icosaedrico	0,20 × 0,09 μm	1 DNA de	57 μm	170	kb
<i>Escherichia coli</i>	Cilindrico	1,7 × 0,65 μm	1 DNA de	1,5 mm	4,2 × 10 ³	kb
Nucleo cellula umana	Sferico	6,00 × 6,00 μm	46 crom., DNA de	2,2 m	6 × 10 ⁶	kb

* sf = singolo filamento; de = doppia elica.

pie originatesi con la replicazione del DNA. Un altro aspetto interessante spesso trascurato è che se si calcola la concentrazione del materiale genetico nel compartimento che lo contiene, ci si rende conto che questa è estremamente elevata, dell'ordine delle decine, fino a centinaia di mg/mL, il che equivale a un gel a elevata viscosità in cui relativamente poche molecole di acqua si infilano negli spazi tra le macromolecole. Questo ha probabilmente delle importanti implicazioni, non ancora affrontate, sui processi di replicazione e di trascrizione.

Andiamo ora ad analizzare più in dettaglio il compattamento dei genomi virali, procariotici ed eucariotici.

4.1 Impacchettamento dei genomi virali

Anche i virus, nella loro apparente semplicità, presentano un problema di organizzazione e compattamento del genoma all'interno del capsido protettivo: il genoma, a seconda dei casi, può essere costituito da DNA o RNA, e questi possono essere a doppio o a singolo filamento.

I capsidi protettivi dei virus sono costituiti da subunità più o meno numerose di una sola o di poche proteine. A questa struttura di base del capsido si possono trovare associate altre proteine, a costituire strutture coinvolte, per esempio, nel riconoscimento e infezione della cellula ospite. Grandezza e forma delle particelle virali rispondono a criteri di economia, e infatti il volume interno del capsido è in genere poco più che sufficiente a contenere il genoma. Dal punto di vista geometrico, ci sono due modi in cui subunità proteiche uguali possono delimitare uno spazio chiuso: (1) le subunità proteiche possono aggregarsi una dietro l'altra in una disposizione elicoidale tipica della struttura dei virus filamentosi; (2) le subunità possono aggregarsi a costituire una struttura cava più o meno sferica tipica della struttura dei virus icosaedrici.

Prenderemo qui in considerazione due casi rappresentativi dell'uno e dell'altro tipo di struttura.

1. Nei **virus filamentosi** il rivestimento proteico si assembla intorno all'acido nucleico organizzandolo in una struttura elicoidale. Il caso più studiato è quello del TMV (*Tobacco Mosaic Virus*, virus del mosaico del tabacco). La struttura tridimensionale di questo virus è stata risolta grazie al contributo fondamentale di Aaron Klug (**fig. 4.1**) combinando analisi di diffrazione dei raggi X e di microscopia elettronica, metodologie che poi ha anche utilizzato per lo studio della struttura del nucleosoma (*vedi* par. 4.8). Il genoma di TMV è costituito da una molecola di RNA a singolo filamento. Su questa molecola di RNA si vanno assemblando, mediante interazioni proteina-RNA e in maniera progressiva, subunità proteiche identiche (**fig. 4.2**). L'assemblaggio non inizia a caso,

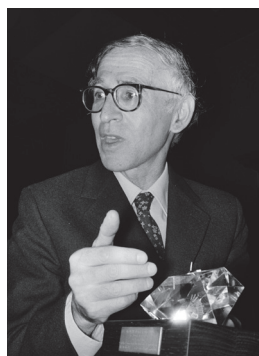
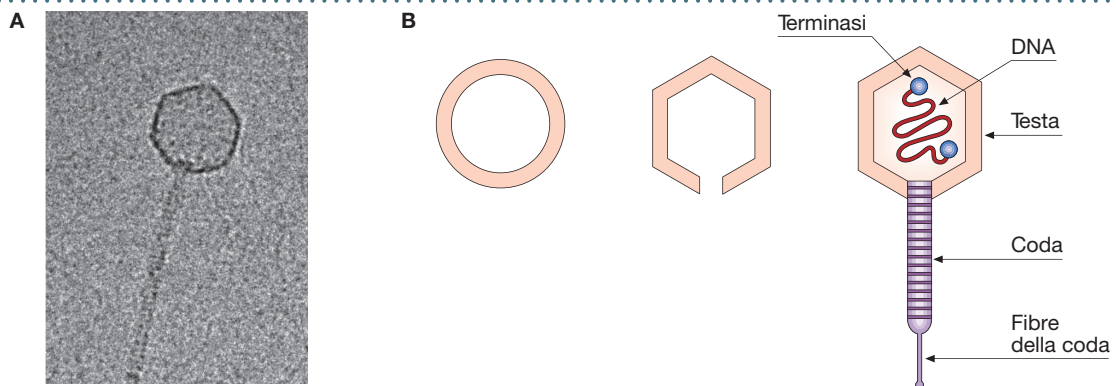
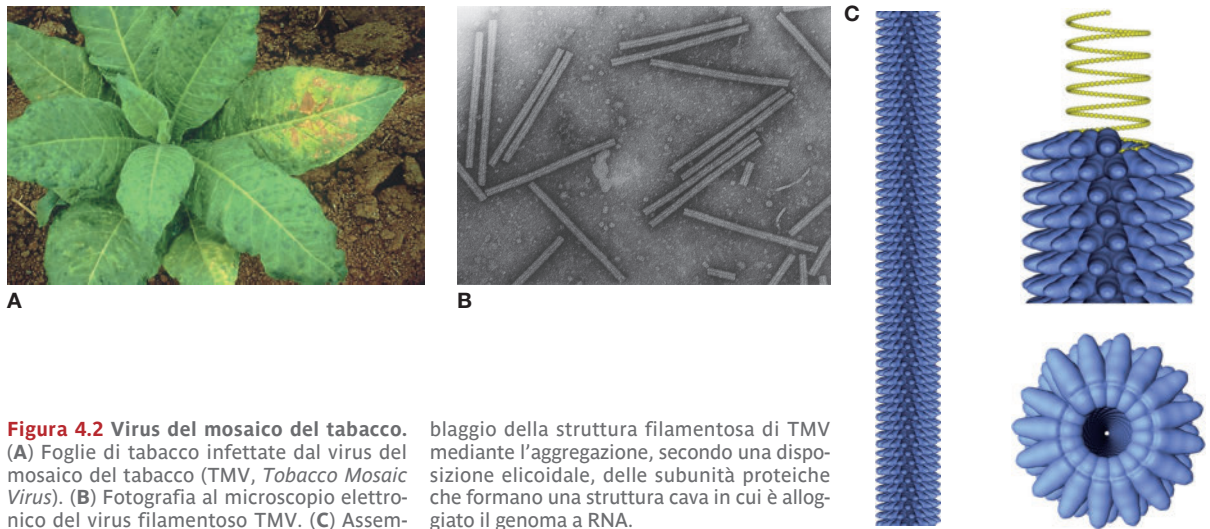


Figura 4.1 Aaron Klug. Premio Nobel per la chimica nel 1982, ha dato un contributo fondamentale allo studio della struttura tridimensionale delle macromolecole biologiche combinando analisi di diffrazione dei raggi X e di microscopia elettronica.

ma in corrispondenza di una specifica struttura a forcina dell'RNA che funge da centro di nucleazione. A partire da questo punto il processo si estende in ambedue le direzioni fino a raggiungere l'estremità della molecola. Durante l'assemblaggio le subunità proteiche si dispongono a formare un'elica, con 17 subunità per ogni giro. L'RNA si avvolge con un andamento elicoidale alla faccia interna dell'involucro proteico che si va formando.

2. I **virus icosaedrici** vengono assemblati in modo diverso. In questo caso, infatti, viene prima assemblato un involucro vuoto nel quale, successivamente, viene introdotto il genoma con un processo che viene chiamato di **impacchettamento** (*packaging*). I virus meglio caratterizzati da questo punto di vista sono il fago T4 e il fago lambda. Il genoma di quest'ultimo è costituito da una molecola di DNA a doppia elica che si trova nella cellula ospite (*E. coli*) in forma circolare. La sua replicazione, mediante il meccanismo del cerchio rotante (*vedi* finestre 6.5 e 8.2), risulta in lunghe molecole di DNA in cui il genoma è ripetuto numerose volte in tandem. Il processo di assemblaggio del fago (**fig. 4.3**) inizia con la formazione di un piccolo involucro contenente un nucleo proteico, che si converte poi in una testa vuota. Solo a questo punto il DNA comincia a essere introdotto nella testa fino a riempirla (processo di **traslocazione**). Il DNA viene quindi tagliato in corrispondenza di siti speci-



ci, chiamati **siti *cos***, che si trovano ripetuti a ogni unità genomica. In questo modo vengono generati genomi lineari completi, con estremità appiccicose corrispondenti ai siti *cos*, che subiscono all'interno dei capsidi un processo di **condensazione**. Va notato che a differenza dei genomi a singolo filamento, il DNA a doppia elica presenta una certa rigidità strutturale che rende sicuramente più problematiche le reazioni di traslocazione e di condensazione. Non a caso, quindi, la traslocazione è un processo attivo, ATP-dipendente. Non è chiaro ancora se il DNA viene "spinto" nella testa dalle proteine coinvolte nel taglio del DNA, o se viene "tirato" dentro da proteine "interne" dei capsidi. Va notato che l'impacchettamento del DNA del fago lambda non è un processo sequenza-specifico, nel senso che non richiede particolari sequenze nucleotidiche nel DNA. Infatti, delezioni, inserzioni e sostituzioni nel genoma di lambda non interferiscono con il processo di impacchettamento; è sufficiente la presenza delle sequenze *cos*, alla giusta distanza di circa 50 kpb, necessarie per il taglio delle unità genomiche.

I due tipi di impacchettamento di genomi virali che abbiamo descritto sono di gran lunga i più frequenti. Esistono tuttavia numerosi casi particolari che si differenziano per un aspetto o per un altro dai due esempi qui descritti.

4.2 Impacchettamento del genoma batterico

I genomi procariotici sono in genere (ma non sempre) costituiti da un singolo cromosoma, come è il caso in *E. coli*. Pur non essendo contenuti all'interno di una ben definita membrana nucleare, questi genomi si trovano in forma di uno o più ammassi abbastanza compatti, chiamati **nucleoidi**, all'interno della cellula, di cui occupano circa un terzo del volume, e appaiono ancorati in un punto alla superficie interna della membrana cellulare (mesosoma). Questo ancoraggio alla membrana è strumentale alla segregazione dei due genomi figli dopo la replicazione, funzione che nelle cellule eucariotiche è svolta dal fuso mitotico, tipicamente assente nei procarioti. Infatti, dopo la replicazione della singola molecola di DNA che costituisce il genoma di *E. coli* i cromosomi figli, ambedue ancorati alla membrana, segregano per effetto della crescita della membrana stessa.

Il nucleotide di *E. coli* è composto per l'80% di DNA, essendo il restante 20% costituito da proteine e RNA. Oltre alle proteine coinvolte più o meno transientemente nei processi di replicazione e trascrizione, sono state isolate varie proteine che si trovano legate al DNA di *E. coli*, e che presentano una qualche somiglianza con le proteine istoniche implicate nell'organizzazione cromatinica delle cellule eucariotiche. Il ruolo di alcune di queste proteine (HU, IHF, H1, P) è stato studiato sia analizzando le loro interazioni con il DNA (*vedi* cap. 20), sia mutandone i geni e andando ad analizzarne i fenotipi correlati. Almeno alcune di queste proteine, come HU e H1, avrebbero un ruolo nella condensazione del nucleotide e nella struttura topologica del DNA. Ci si sarebbe aspettato che alterazioni di proteine coinvolte nell'organizzazione e nel compattamento del genoma determinassero una generale disorganizzazione cromosomica e, quindi, fenotipi letali o comunque molto gravi. Si sono osservati, invece, effetti piuttosto blandi suggerendo che forse queste diverse proteine hanno probabilmente funzioni simili e in parte sovrapposte. In sostanza, un effetto di ridondanza genica attenuerebbe l'effetto causato da mutazioni di singoli geni.

L'intero cromosoma di *E. coli* è costituito da un'unica molecola circolare di DNA organizzata in circa un centinaio di anse della lunghezza media di circa 40 kpb (**fig. 4.4**). Ciascuna ansa è ancorata alla base a una struttura non ben definita, costituendo così, dal punto di vista topologico, un dominio indipendente. Infatti, questo ancoraggio impedisce la propagazione da un dominio all'altro di eventi di rotazione della molecola di DNA, consentendo a regioni diverse del cromosoma

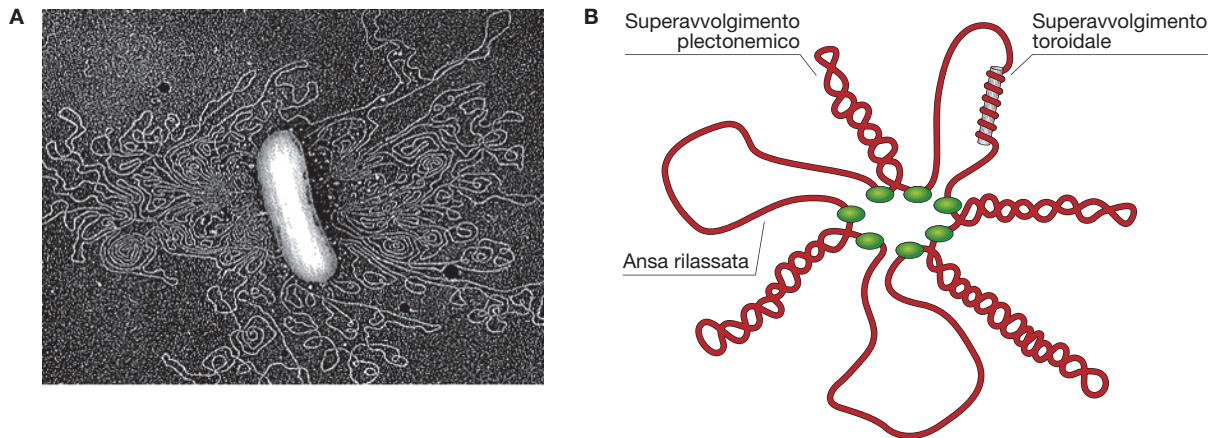


Figura 4.4 Struttura del nucleotide di *E. coli*. (A) Immagine al microscopio elettronico del nucleotide che fuoriesce da una cellula lisata di *E. coli*. (B) Nel nucleotide il DNA circolare è organizzato in forma di un centinaio di anse della lunghezza media di 40 kpb, an-

corate a complessi proteici ancora non ben caratterizzati (in verde) che le rendono topologicamente indipendenti. Si possono osservare anse rilassate o superavvolte in forma plectonemica o toroidale (vedi anche par. 2.3).

di mantenere, per azione di girasi e topoisomerasi deputate al mantenimento degli equilibri topologici, diversi gradi di superavvolgimento con conseguenti effetti sulla regolazione dell'espressione genica.

Va notato come le nostre conoscenze sulla struttura e organizzazione del cromosoma batterico nel nucleotide siano ancora molto incomplete. Come vedremo qui appresso, sappiamo di più sull'organizzazione cromatinica e nucleosomiale dei cromosomi eucariotici.

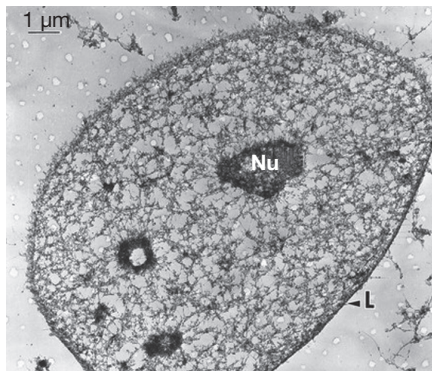
4.3 Organizzazione e impacchettamento del DNA eucariotico

Nei nuclei delle cellule eucariotiche il materiale genetico costituisce la massa di cromatina la cui organizzazione è soggetta a vistosi cambiamenti nel corso del ciclo cellulare. Durante l'interfase, che rappresenta la maggior parte del ciclo cellulare quando il DNA viene trascritto e replicato, le fibre cromatiniche si trovano disperse nel nucleo a occuparne la maggior parte del volume. Durante la più breve fase mitotica, quando il DNA ormai replicato deve essere segregato nelle due cellule figlie, si assiste a un ulteriore compattamento del genoma a costituire i tipici cromosomi mitotici ben distinguibili al microscopio ottico.

Le molecole di DNA che costituiscono i cromosomi in una cellula eucariotica interfase sono associate alle proteine istoniche a formare, al più basso livello di organizzazione, la fibra cromatinica da 10 nm di diametro. Queste lunghissime molecole di DNA sono organizzate, analogamente a come abbiamo visto per il nucleotide batterico, in numerose anse della lunghezza media di 30-100 kpb ancorate alla base su una struttura filamentosa chiamata **matrice nucleare** (fig. 4.5). Come abbiamo già visto nel caso del nucleotide batterico, queste anse costituiscono domini topologici indipendenti in quanto capaci individualmente di mantenere o perdere superavvolgimenti.

È interessante che anche nei cromosomi mitotici, quando il DNA si trova nello stato di massimo compattamento presenta un'organizzazione con anse di circa 30-100 kpb. Queste anse sono ancorate a una "impalcatura" centrale del cromosoma costituita da proteine specifiche. È probabile che ci sia una relazione diretta tra le anse di DNA del nucleo interfase e quelle dei cromosomi mitotici, e forse anche tra "matrice nucleare" interfase e "impalcatura" del cromosoma mitotico.

Figura 4.5 Matrice nucleare di una cellula umana visualizzata al microscopio elettronico. La matrice nucleare è costituita da due parti, la lamina nucleare (L) e un'intricata rete di fibre connesse alla lamina, ben distribuite nel volume nucleare. In questa sezione si possono osservare tre nucleoli (Nu) connessi alle fibre della matrice nucleare. [Fonte: Nickerson, J.A. et al. (1997), *Proc. Natl. Acad. Sci. USA*, **94**, pp. 4446-4450.]



Una domanda che ci si è posti è se nel DNA esistano delle sequenze specifiche per l'ancoraggio alla matrice nucleare e all'impalcatura cromosomica. Una serie di esperimenti ha permesso di isolare e analizzare queste sequenze di ancoraggio e di concludere che le stesse sequenze sono implicate nell'ancoraggio sia alla matrice che all'impalcatura. Tali sequenze vengono in genere chiamate **MAR** (*Matrix Attachment Region*), riferendosi alle regioni che si legano alla matrice nucleare durante l'interfase, o **SAR** (*Scaffold Attachment Region*) riferendosi alle sequenze che si legano all'impalcatura dei cromosomi mitotici. Il confronto delle sequenze nucleotidiche di numerose MAR non ha permesso di individuare alcuna specifica sequenza conservata. Tuttavia, sono state notate alcune caratteristiche comuni: in particolare, le regioni contenenti le MAR sono spesso ricche di A-T e sono adiacenti a tratti di DNA intrinsecamente "curvo" (vedi cap. 2). Inoltre, si è notato che spesso, nelle adiacenze delle MAR, ci sono altri rilevanti siti regolativi, come siti di riconoscimento per la topoisomerasi II e siti *cis*-agenti per la regolazione della trascrizione.

4.4 Cromatina interfasica e cromosomi mitotici

La fibra cromatinica risulta dall'interazione del DNA genomico con le proteine istoniche e subisce notevoli variazioni della compattazione nel corso del ciclo cellulare. Durante l'interfase (fasi G1, S e G2 del ciclo), quando avvengono trascrizione e replicazione del DNA, la cromatina si trova dispersa nel nucleo. Nella maggior parte dello spazio nucleare le fibre cromatiniche appaiono relativamente disperse, molto meno densamente compattate che nei cromosomi mitotici, che in questa fase non sono distinguibili, e prendono il nome di **euromatina**. In alcune regioni nucleari invece si osservano delle masse di cromatina più compatta, quasi come nei cromosomi mitotici, che prendono il nome di **eterocromatina**. Quindi euromatina ed eterocromatina rappresentano gradi diversi di condensazione del materiale genetico. L'eterocromatina interfasica può essere distinta in due tipi:

1. l'**eterocromatina costitutiva** resta sempre compatta e rappresenta regioni del genoma che, non avendo capacità codificante, non vengono mai espresse e che potrebbero avere un ruolo strutturale nel cromosoma; essa comprende, per esempio, alcuni DNA satelliti che spesso si trovano localizzati nelle regioni centromeriche;
2. l'**eterocromatina facoltativa**, che è tale solo in alcune situazioni, rappresenta regioni che hanno capacità codificante che si esplica però solo in alcune condizioni: per esempio, regioni non espresse in un tessuto ma che possono essere espresse in un altro tessuto; oppure il ben noto caso del cromosoma X dei mammiferi, di cui una delle due copie presenti nelle cellule delle femmine viene mantenuta inattiva sotto forma di eterocromatina (corpo di Barr) mentre l'altra copia euromatica si esprime normalmente.



Figura 4.6 Bandeggio G (Giemsa) di un corredo cromosomico umano. Questa tecnica di bandeggio cromosomico, ottenuto mediante colorazione di Giemsa, genera in ciascun cromosoma una serie caratteristica di bande trasversali.

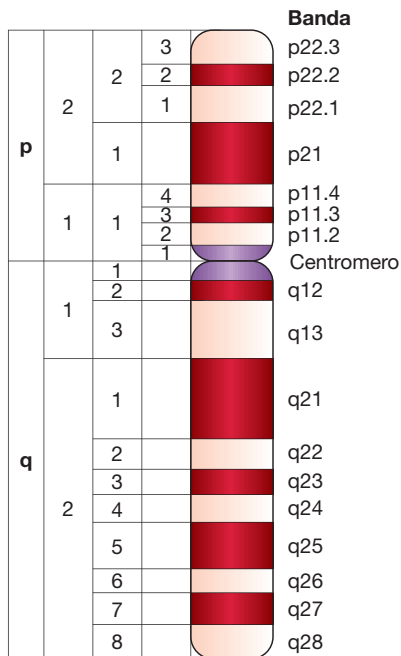


Figura 4.7 Struttura citogenetica del cromosoma X umano. Sulla base del bandeggio G si possono distinguere, nel cromosoma X umano, diverse bande e sottobande sia nel braccio corto p sia nel braccio lungo q.

Il massimo grado di compattamento si realizza nei cromosomi mitotici che così diventano visibili e distinguibili solo per breve tempo al microscopio ottico, quando devono segregare nelle cellule figlie (fase M del ciclo cellulare). Ogni cromosoma contiene un'unica molecola di DNA a doppia elica, in genere molto lunga, che corre da un estremo all'altro. Poiché durante la fase S il DNA è stato replicato, ogni cromosoma mitotico appare come due filamenti (cromatidi fratelli) tenuti insieme nella regione del centromero. Ogni cromosoma del corredo cromosomico di un organismo presenta una struttura diversa e riproducibile, innanzitutto per grandezza e per posizione del centromero. Inoltre l'introduzione di particolari tecniche di colorazione (ad es. il Giemsa) permette di distinguere, in base alla distribuzione delle bande di colorazione, anche cromosomi molto simili per grandezza e posizione del centromero, come mostrato in **figura 4.6** per i cromosomi del genoma umano. Queste tecniche di "bandeggio" sono abbastanza riproducibili così da permettere non solo l'individuazione di ogni singolo cromosoma, ma anche l'identificazione di traslocazioni da un cromosoma a un altro, purché implicino un lungo tratto genomico. Infatti, le bande visibili sono

strutture molto grandi che contengono tratti di DNA lunghi milioni di pb e quindi centinaia di geni. La **figura 4.7** mostra la schematizzazione della struttura citogenetica di un cromosoma.

4.5 Organizzazioni atipiche di cromosomi e genomi

In alcune specie e in alcuni particolari tipi di cellule si trovano strutture e organizzazioni di cromosomi e genomi che si discostano da quelle tipiche sopra descritte. Consideriamo qui alcune di queste situazioni.

Cromosomi a spazzola

Chiamati talvolta anche **cromosomi piumosi**, si possono trovare in cellule di vari organismi, ma sono stati studiati soprattutto quelli che si trovano nel nucleo degli oociti degli anfibi, dove sono visibili durante il lungo processo meiotico, che può durare anche qualche mese. Durante questo periodo i cromosomi si trovano in una forma molto distesa e possono essere facilmente osservati al microscopio ottico. Essendo in fase meiotica, i cromosomi a spazzola sono coppie bivalenti meiotiche, cioè sono costituiti ciascuno da due cromosomi omologhi, tenuti uniti in pochi punti tramite i “chiasmi”. Ognuno dei due cromosomi omologhi, costituito a sua volta da una coppia di cromatidi fratelli, appare al microscopio ottico come un asse centrale da cui sporgono lateralmente delle anse. Queste anse sono continue con il filamento assiale e rappresentano materiale cromosomico estruso dall'organizzazione più compatta presente nel **cromomero** (**fig. 4.8**). Le anse sono circondate da una matrice ribonucleoproteica, nella quale si trovano le catene nascenti di RNA. Ogni ansa rappresenta quindi una regione di DNA estruso che viene attivamente trascritto. Proprio nelle stesse cellule, copie extracromosomiche dei geni ripetuti in tandem per l'rRNA sono anche trascritte attivamente e appaiono al microscopio elettronico come caratteristiche strutture ad “albero di Natale” (*vedi fig. 12.3*).

Cromosomi politenici

La regolazione del ciclo cellulare (*vedi capp. 6 e 7*) garantisce in genere che ci sia una precisa alternanza tra replicazione dei cromosomi e segregazione nelle due cellule figlie dei cromatidi fratelli così originati: in altre parole, non si può avere mitosi se non è stata completata la replicazione dei cromosomi durante la fase S, e non si può avere replicazione dei cromosomi se non è stata completata la segregazione mitotica. A questa regola generale esistono però delle eccezioni, la più notevole delle quali è la **politenia**. Questo processo è dovuto al susseguirsi di varie replicazioni cromosomiche non seguite da segregazione, in cui i cromatidi fratelli originati dalle replicazioni rimangono strettamente affiancati a formare dei fasci di fibre parallele che rappresentano, appunto, i cromosomi politenici. Fenomeni di politenia avvengono in vari tipi di cellule differenziate di numerosi organismi, ma sicuramente i cromosomi politenici più studiati sono i **cromosomi giganti** che si trovano nelle ghiandole salivari di *Drosophila melanogaster* (*vedi anche par. 6.8*). Come si vede in **figura 4.8**, ogni cromosoma contiene una serie di bande la cui larghezza varia da 0,05 a 0,5 μm . Le bande contengono la gran parte del DNA e si colorano con opportuni reagenti, mentre le regioni che si alternano con le bande, chiamate **interbande**, si colorano meno intensamente. Nei cromosomi politenici di *Drosophila* si contano circa 5000 bande. Il normale patrimonio genetico di *Drosophila* è costituito da quattro paia di cromosomi omologhi. Nei cromosomi politenici ciascuna coppia di omologhi è appaiata a formare un'unica struttura che per successivi cicli di replicazione darà origine a un cromosoma politenico. I quattro cromosomi politenici così generati sono uniti tra loro al livello dei centromeri che costituiscono una massa unica da cui dipartono tutte

le braccia dei cromosomi politenici. Nella figura si può notare che lo spessore del fascio di fibre non è costante lungo il cromosoma, presentando tratti con diametro maggiore e tratti con diametro minore. Ciò è dovuto al fatto che il numero di repliche avvenuto non è esattamente lo stesso lungo tutto il cromosoma. In genere si hanno al massimo 10 repliche, che risultano nella formazione di circa mille fibre parallele ($2^{10} = 1024$), ma in alcune regioni una o due repliche di meno risulteranno nella metà o $\frac{1}{4}$ del numero di fibre parallele. Mentre nelle ghiandole salivari la politenizzazione è più o meno simile lungo tutti i cromosomi, in altri tessuti essa è molto più variabile, risultando in cromosomi politenici che presentano ingrossamenti e costrizioni molto più evidenti. Dal punto di vista della replicazione, la politenizzazione è dovuta ad attivazioni successive di origini di replicazione del DNA senza che queste repliche vengano poi portate tutte a termine (vedi anche par. 6.8 e fig. 6.33).

Quando lungo un cromosoma politenico un certo gene, o meglio le numerose

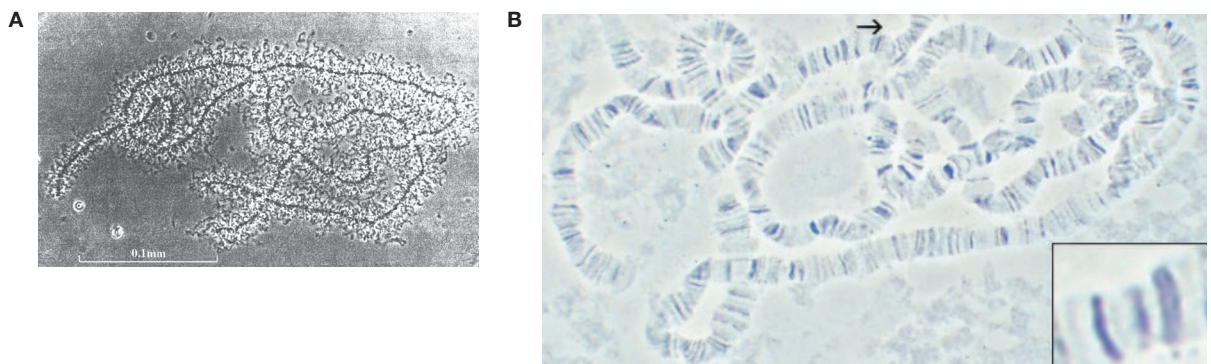


Figura 4.8 Cromosomi a spazzola e cromosomi politenici. Fotografie al microscopio ottico di (A) cromosomi a spazzola nel nucleo di un oocita di anfibio e (B) di un tratto di cromosoma politenico di *Drosophila*.

(Fonte: http://www.natureprotocols.com/2007/05/09/highresolution_light_microscop.php).

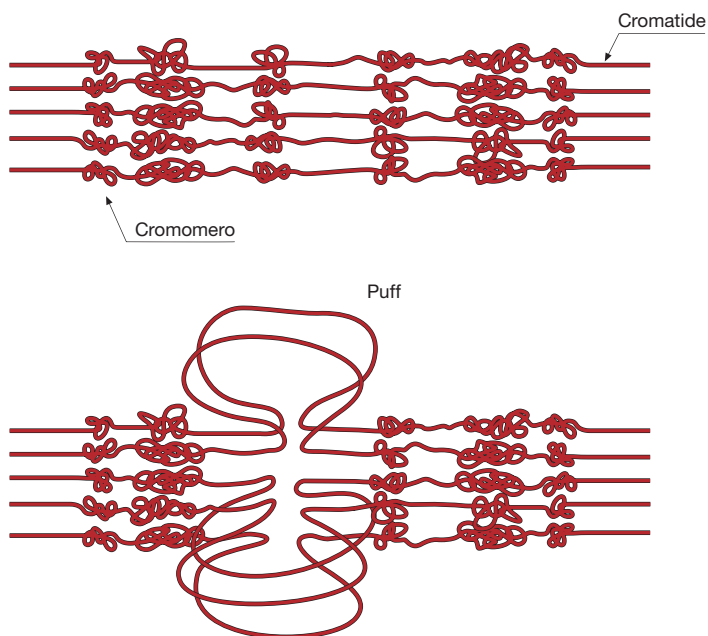


Figura 4.9 Formazione di estrusioni di regioni di DNA trascrizionalmente attive in cromosomi politenici, denominate "puff".

copie parallele di quel gene, viene attivato trascrizionalmente, i tratti di DNA trascritti vengono estrusi dal cromosoma in forma di anse che, nel loro insieme, formano un **puff**, che rappresenta una struttura visibile al microscopio ottico dell'attività di un gene (**fig. 4.9**).

Micronucleo e macronucleo dei ciliati

Un'organizzazione molto particolare del genoma e dei cromosomi è quella che caratterizza i protozoi ciliati quali *Tetrahymena* o *Paramecium*. In tali organismi sono presenti due nuclei distinti, che differiscono sia nella forma che nella funzione, ma contengono lo stesso corredo di geni. Il micronucleo diploide, presente in due copie in *P. tetraurelia*, ha una funzione germinale ed è completamente inattivo per quanto concerne la trascrizione. La struttura, la funzione e le modalità di trasmissione del macronucleo e micronucleo dei ciliati sono descritte nel capitolo 6 (*vedi par. 6.8*).

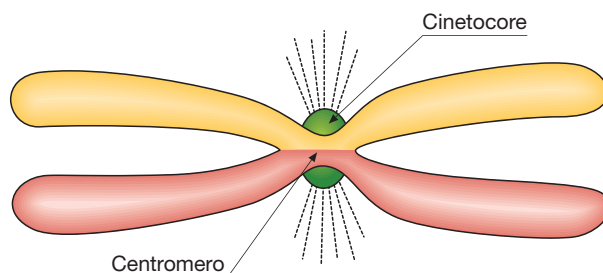
4.6 Centromeri

Ogni cromosoma è dotato di un centromero, struttura essenziale per la corretta segregazione dei due cromatidi fratelli nelle cellule figlie. Il centromero è costituito da un complesso proteico assemblato su una specifica sequenza di DNA. A esso si lega l'estremità di un microtubulo, mentre l'altra estremità è connessa a un polo del fuso mitotico permettendo la migrazione dei cromosomi durante l'anafase del processo mitotico. All'interno della regione cromosomica del centromero si può distinguere il **cinetocore**, struttura fibrosa a cui è direttamente attaccato il microtubulo (**fig. 4.10**).

Le sequenze di DNA centromerico presentano, nei diversi organismi eucariotici, una enorme variabilità di grandezza e di complessità, dai piccoli centromeri di circa 120 pb del lievito *Saccharomyces cerevisiae* fino alle milioni di paia di basi di DNA ripetitivo dei centromeri dei vertebrati.

I centromeri di lievito, anche a causa della loro relativa semplicità, sono stati i primi a essere isolati sfruttando la loro capacità di conferire una corretta segregazione a un cromosoma. L'isolamento e il sequenziamento delle regioni centromeriche (chiamate **CEN**) dei sedici cromosomi di *Saccharomyces cerevisiae* hanno permesso di identificare una struttura comune. Come mostrato in **figura 4.11**, una regione CEN, che è lunga circa 120 pb, è costituita da tre elementi: l'elemento centrale CDE-II, fiancheggiato dai due elementi CDE-I e CDE-III. CDE-II è lungo circa 90 pb e, pur non avendo una sequenza conservata nei vari centromeri, è comunque sempre molto ricco (oltre 90%) di coppie di basi A·T; CDE-I e CDE-III sono brevi sequenze, rispettivamente di 9 e 11 pb, molto conservate nei diversi centromeri. Il centromero è avvolto intorno a un complesso proteico che comprende varie proteine, alcune delle quali collegano il centromero al microtubulo. Tra le proteine centromeriche si possono ricordare la proteina Cse4p, che

Figura 4.10 Il centromero. A livello del centromero, che assicura la corretta segregazione del DNA durante la mitosi, si assembla uno specifico complesso proteico denominato "cinetocore", cui è legata una delle estremità dei microtubuli.



A

TCACATGATGATATTTGATTTTATTATATTTTTAAAAAGTAAAAATAAAAAGTAGTTTATTTTTAAAAATAAAAATTTTCAAAAAATGATTTCCGAA
 AGTGTACTACTATAAACTAAATAATATAAAATTTTTTCATTTTTATTTTCATCAATAAAAAATTTTTATTTTAAATTTTATAAAGTGTTTTACTAAAGGCTT

CDE-I CDE-II CDE-III

B

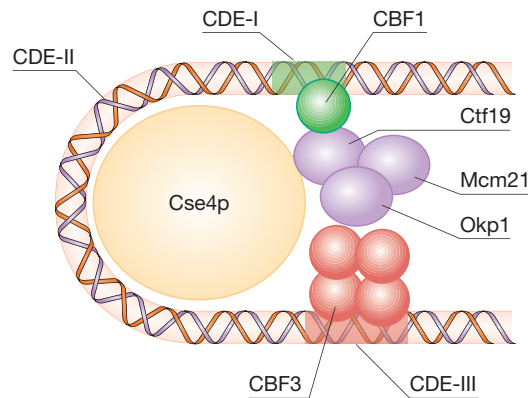


Figura 4.11 Regione CEN del centromero di lievito. Tale regione è costituita dai domini CDE-I, CDE-II e CDE-III (A) ed è avvolta intorno a un complesso proteico (B).

interagisce con CDE-II, e le proteine CBF1 e CBF3, che interagiscono, rispettivamente, con CDE-I e CDE-III. Le restanti proteine sono verosimilmente coinvolte nel legame ai microtubuli e implicate nella segregazione dei cromosomi.

Nella maggior parte delle altre specie studiate i centromeri appaiono alquanto più complessi rispetto ai semplici centromeri di *S. cerevisiae*, a causa della presenza di sequenze di DNA altamente ripetitivo (DNA satelliti) nei centromeri stessi e/o nelle loro vicinanze, dove costituiscono zone di eterocromatina costitutiva. Già i centromeri del lievito *Schizosaccharomyces pombe* sono alquanto complessi essendo stati localizzati in regioni di 50-100 kpb costituite, sostanzialmente, da DNA ripetitivo. Nei cromosomi di *Drosophila* la funzione centromerica sembra dispersa su lunghe regioni di circa 200-600 kpb. Nei primati, il componente principale del centromero è costituito da DNA satellite, cioè una sequenza semplice ripetuta moltissime volte in tandem. In particolare, nel topo una sequenza di 234 pb, evolutasi per una serie di duplicazioni e mutazioni a partire da una breve sequenza di 9 pb (vedi cap. 5, fig. 5.17), è a sua volta ripetuta molte volte in tandem nelle regioni centromeriche di tutti i cromosomi. Nell'uomo i centromeri sono costituiti da un DNA satellite (α -satellite), che consiste in una serie di ripetizioni in tandem non perfette di una unità di 171 pb che si estendono per lunghezze di Mpb. Non è chiaro se siano le stesse sequenze del DNA satellite a svolgere la funzione centromerica o piuttosto altre sequenze intramezzate con esso. È stato proposto anche che, nel corso dell'evoluzione, i centromeri, di per sé relativamente semplici, siano un sito bersaglio per inserzioni di DNA ripetitivo.

4.7 Telomeri

Le estremità delle molecole lineari di DNA sono instabili nel tempo e nelle generazioni per vari motivi tra cui:

1. degradazione da parte di esonucleasi;
2. impossibilità, che si viene a determinare copiando l'estremità 3' di un filamento di DNA, di replicare i primi nucleotidi per mancanza di un innesco (vedi fig. 6.16);

3. propensione a legare altre molecole di DNA dato che le estremità rotte di cromosomi sono “appiccicose”;
4. attivazione dei sistemi di controllo che riconoscono rotture nel DNA e che eventualmente attivano l'apoptosi (*vedi* cap. 7).

L'evoluzione ha escogitato vari meccanismi per ovviare a questi problemi di instabilità. In molti casi, semplicemente, i cromosomi non hanno estremità, cioè sono molecole di DNA circolari, come è il caso dei cromosomi di molti virus, batteri, plasmidi, cloroplasti, mitocondri ecc. In alcuni casi, non frequenti, la molecola lineare di DNA presenta delle sequenze uguali ripetute ai due estremi. Infine, particolari sequenze specializzate, chiamate **telomeri**, stabilizzano le due estremità della lunghissima sequenza di DNA lineare che costituisce ciascun cromosoma eucariotico. Le prime a essere identificate e studiate sono state, per motivi tecnici, le sequenze di DNA telomerico del protozoo ciliato *Tetrahymena*. Come abbiamo già visto, questo protozoo presenta due tipi di nucleo, il micronucleo, che contiene l'intero genoma suddiviso in cinque coppie di cromosomi, utilizzato per la trasmissione dell'informazione genetica alla progenie, e il macronucleo, in cui le cinque coppie di cromosomi sono frammentate in oltre 200 minicromosomi, utilizzati per l'espressione genica, dotati alle estremità di sequenze telomeriche. La relativa abbondanza di telomeri facilitò il clonaggio e il sequenziamento delle sequenze telomeriche e l'individuazione dell'attività enzimatica della telomerasi (*vedi* anche par. 6.6). Studi successivi su piante e animali superiori, incluso l'uomo, hanno mostrato comunque un'organizzazione molto simile, indicando che l'organizzazione strutturale dei telomeri è simile in diversi organismi. Il DNA telomerico è in genere costituito da centinaia di copie di una sequenza ripetuta in tandem di 5-8 pb (in **tabella 4.2** sono riportate le unità ripetitive identificate per telomeri di varie specie). Il meccanismo di formazione dei telomeri ad opera della telomerasi è illustrato in figura 6.24, in cui si vede che il filamento con l'estremità 3' è ricco nelle basi G+T e protrude rispetto al filamento complementare ricco di C+A.

Tabella 4.2 Unità ripetute delle sequenze telomeriche in diversi organismi.

Tipo di organismo	Nome scientifico	Ripetizione telomerica (direzione 5'→3')
Vertebrati	<i>Homo sapiens</i> , <i>Mus musculus</i> , <i>Xenopus laevis</i>	TTAGGG
Funghi	<i>Neurospora crassa</i> , <i>Physarum</i> , <i>Didymium</i>	TTAGGG
Protisti	<i>Dictyostelium discoideum</i>	AG(1-8)
Kinetoplastea (protozoi)	<i>Trypanosoma</i> , <i>Crithidia</i>	TTAGGG
Protozoi ciliati	<i>Tetrahymena</i> , <i>Glaucoma</i>	TTGGGG
	<i>Paramecium</i>	TTGGG(T/G)
	<i>Oxytricha</i> , <i>Stylonychia</i> , <i>Euplotes</i>	TTTTGGGG
Apicomplexa	<i>Plasmodium</i>	TTAGGG(T/G)
Piante superiori	<i>Arabidopsis thaliana</i>	TTTAGGG
Alghe verdi	<i>Chlamydomonas</i>	TTTTAGG
Insetti	<i>Bombyx mori</i>	TTAGG
Anellidi	<i>Ascaris lumbricoides</i>	TTAGGC
Lieviti a scissione binaria	<i>Schizosaccharomyces pombe</i>	TTAC(A)(C)G(1-8)
Lieviti gemmanti	<i>Saccharomyces cerevisiae</i>	TGTGGGTGTGGTG
	<i>Candida glabrata</i>	GGGGTCTGGGTGCTG

La lunghezza del telomero non è costante, a causa di variazioni del numero di ripetizioni. Il progressivo accorciamento del telomero, dovuto ai problemi di replicazione delle estremità del DNA, viene compensato da un periodico allungamento mediante addizione di nuove ripetizioni telomeriche. L'allungamento e l'accorciamento del telomero sono quindi in equilibrio dinamico.

Con quale meccanismo i telomeri conferiscono stabilità all'estremità dei cromosomi? In parte, come già detto, mediante la sintesi di nuove unità di ripetizione che ne compensano la costante perdita. Ci sono anche proteine, come TRF1 e TRF2 (TTAGGG *Repeat binding Factor* 1 e 2) nei mammiferi, che riconoscono e legano la sequenza telomerica, proteggendo così le estremità del DNA cromosomico. È stato anche osservato che le estremità dei telomeri non sono lineari, ma assumono una struttura a forma di loop, denominata **T-loop** (*telomere-loop*). Questa struttura, mostrata e descritta in **figura 4.12**, è dovuta all'appaiamento tra le sequenze TTAGGG dell'estremità 3' che protrude a singolo filamento e il filamento complementare in un breve tratto denaturato nella lunga regione ripetitiva adiacente duplex. La struttura è stabilizzata dalla formazione di complessi con numerose proteine implicate nella funzione telomerica e nella protezione dell'estremità del cromosoma. Inoltre, è stata descritta un'altra particolare struttura del DNA che verrebbe assunta dal filamento protruso del telomero e che potrebbe avere un ruolo rilevante nella stabilizzazione. Questa struttura, chiamata **quartetto di G**, è formata da quattro residui G, distanziati lungo un singolo filamento a poche basi uno dall'altro (**fig. 4.13**). A causa di tre ripiegamenti a forcina del filamento, le quattro G si vengono a trovare sullo stesso piano e formano legami idrogeno, come abbiamo già descritto nel capitolo 2 (*vedi* par. 2.3 e fig. 2.24). Poiché il filamento con estremità 3' protrusa del telomero ha sequenze ripetute in tandem contenenti una serie (3 o 4) di G, si possono formare fino a 3 o 4 quartetti di G che, impilandosi l'uno sull'altro, possono dare origine a una struttura molto stabile che fornirebbe una protezione all'estremità del cromosoma.

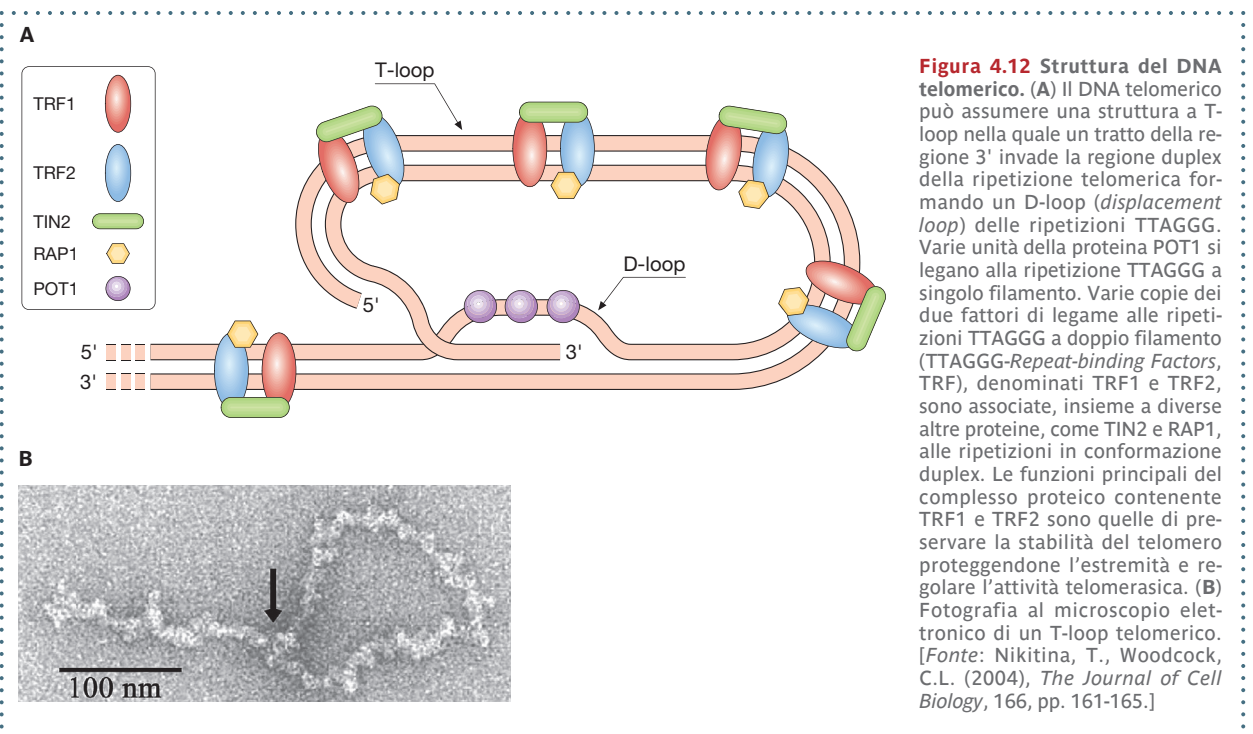
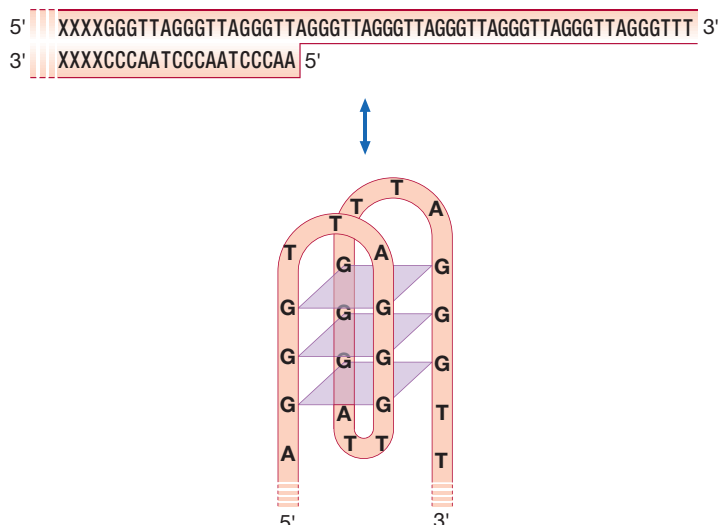


Figura 4.13 Il DNA telomerico può adottare una struttura a quartetti di G. Sopra è mostrata la sequenza nucleotidica dell'estremità a singolo filamento di un telomero umano (vedi tab. 4.2) e, sotto, la struttura a quartetti di G che essa può assumere.



4.8 Nucleosomi e proprietà strutturali e funzionali della cromatina

Come abbiamo già accennato nel paragrafo 4.3, nelle cellule eucariotiche le molecole di DNA che costituiscono i cromosomi sono strettamente associate a proteine, costituendo il complesso nucleoproteico chiamato **cromatina**. Quando, infatti, si lisa il nucleo di una cellula eucariotica, si osserva la fuoriuscita di un ammasso cromatinico costituito dal DNA, dall'RNA che risulta dalla trascrizione dei geni e da proteine di vario tipo. Tra queste, oltre alle numerose e varie proteine più o meno acide che sono gli enzimi e i fattori implicati nei processi di replicazione e trascrizione del DNA, troviamo gli **istoni**, proteine basiche che rappresentano i veri costituenti strutturali della cromatina. Difatti, le proteine della cromatina sono costituite per l'80-90% da istoni, con la restante frazione proteica costituita per lo più da altre piccole proteine basiche non istoniche denominate **High-Mobility Group (HMG)**.

I primi indizi sull'organizzazione strutturale della cromatina sono venuti dalle osservazioni al microscopio elettronico fatte negli anni settanta, indipendentemente, da Donald e Ada Olins (**fig. 4.14**) e da Christopher Woodcock. Questi studi hanno permesso di osservare la struttura ripetitiva della cromatina che appare come una collana costituita da grani impilati su un filo, dove i grani rappresentano i **nucleosomi** (**fig. 4.15**), le unità elementari della struttura della cromatina. Un altro importante contributo per la risoluzione della struttura della cromatina e dei nucleosomi è derivato dagli studi condotti negli anni settanta e ottanta mediante digestione della cromatina con nucleasi di vario tipo. Come mostrato in **figura 4.16A**, se si digerisce la cromatina con la nucleasi micrococca, le dimensioni del DNA tagliato rappresentano multipli interi di un'unità nucleosomiale, pari a circa 200 pb. Se invece si effettua una digestione spinta con nucleasi, fino a digerire completamente il tratto di DNA linker tra due nucleosomi, si ottiene un'unica banda di lunghezza pari a 147 pb che corrisponde al DNA strettamente avvolto al nucleo del nucleosoma (**fig. 4.16B**). Mentre le dimensioni delle unità monomeriche di DNA possono variare da organismo a organismo, e nello stesso organismo in diversi tessuti e tipi cellulari, le dimensioni del tratto di DNA legato al nucleo del nucleosoma sono strettamente conservate. La diversa lunghezza dell'unità monomerica è dunque dovuta alla variabilità delle dimensioni della regione linker.



Figura 4.14 Ada L. Olins e Donald E. Olins hanno ottenuto (insieme a Chris L. Woodcock), mediante studi di microscopia elettronica, le prime informazioni di base sull'organizzazione della cromatina in unità nucleosomali.

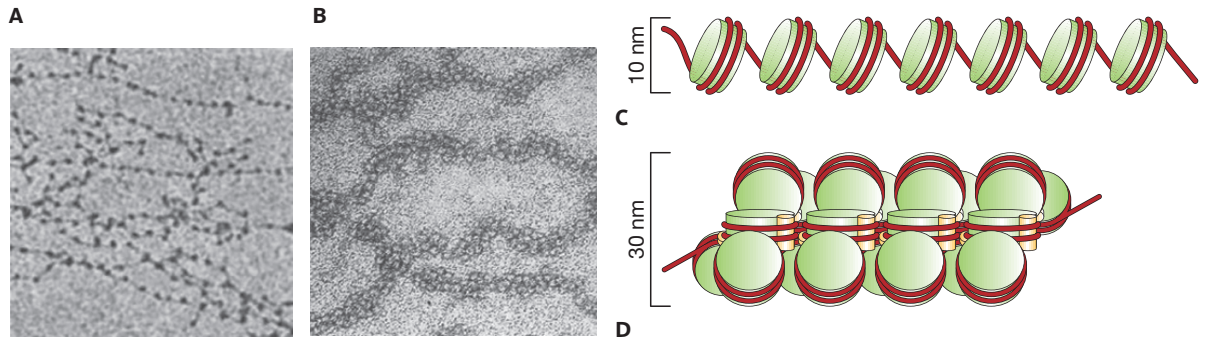


Figura 4.15 Organizzazione del DNA in fibre cromatiniche. (A) In assenza dell'istone H1 la cromatina appare, al microscopio elettronico, come una serie di grani su un filo che costituisce la fibra da 10 nm di diametro, chiamata anche "collana di perle". (B) In presenza di H1 la fibra da 10 nm si organizza in una struttura supe-

riore, chiamata fibra da 30 nm; (C e D) mostrano, rispettivamente, una rappresentazione schematica della fibra da 10 nm e uno dei possibili modelli per la fibra da 30 nm che la descrive come struttura elicoidale ("solenoidale") con sei nucleosomi per giro.

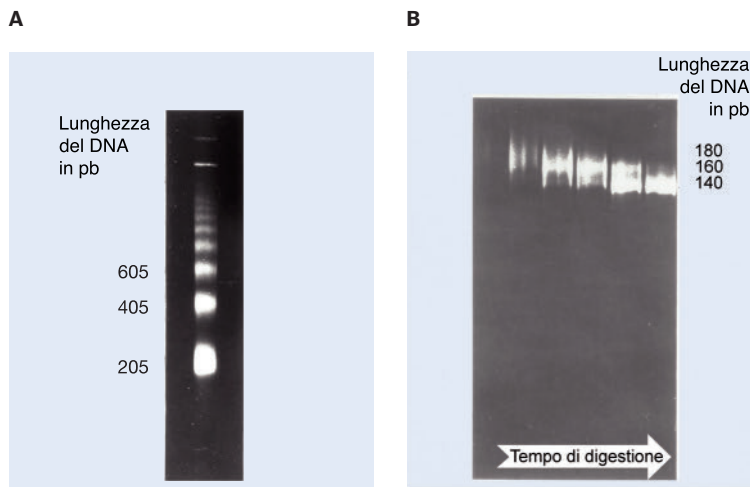


Figura 4.16 Digestione con nucleasi micrococcica del DNA nucleosomiale. (A) La digestione blanda della cromatina con nucleasi micrococcica, agendo casualmente solo su una frazione dei DNA linker, produce una serie di frammenti di dimensioni corrispondenti a multipli di unità nucleosomiali pari a circa 200 pb. (B) Aumentando il tempo di digestione si ottiene una progressiva riduzione delle dimensioni del DNA nucleosomiale fino a circa 147 pb, quando tutto il DNA linker è stato degradato e rimane solo il DNA legato al nucleo del nucleosoma.

A questi studi effettuati tramite microscopia elettronica e mediante digestione della cromatina con nucleasi, sono seguiti studi strutturali più approfonditi mediante diffrazione dei raggi X ottenuta con cristalli di nucleosomi. Tali ricerche, a cui hanno contribuito vari ricercatori tra cui Aaron Klug che abbiamo già menzionato (vedi **fig. 4.1**) e Timothy J. Richmond, hanno portato a modelli strutturali molto dettagliati (**fig. 4.17A**). Tra l'altro si è potuto determinare che la forte interazione tra DNA e core istonico è mediata da circa 140 legami a idrogeno (una tipica interazione DNA-proteina ne coinvolge circa 20). Quasi tutti i legami si instaurano con gli atomi di ossigeno dei legami fosfodiesterici vicini al solco minore; solo 7 legami vengono formati tra le proteine e le basi attraverso il solco maggiore. Da qui la conclusione molto importante che il legame tra DNA e nucleosoma è forte ma non è sequenza-specifico. La natura basica degli istoni maschera le cariche negative dei fosfati, permettendone un avvicinamento fisico, determinato dalla curvatura del DNA, altrimenti impossibile.

Oggi sappiamo che il nucleosoma è costituito da un tratto di DNA di circa 200 pb associato con un ottamero di proteine istoniche, che consiste di due copie di ciascuno degli istoni H2A, H2B, H3 e H4 (**tab. 4.3**). Di questi, H3 e H4 sono tra le proteine evolutivamente più conservate che conosciamo, indicando che svolgono una funzione importante rimasta identica in tutti gli eucarioti. An-

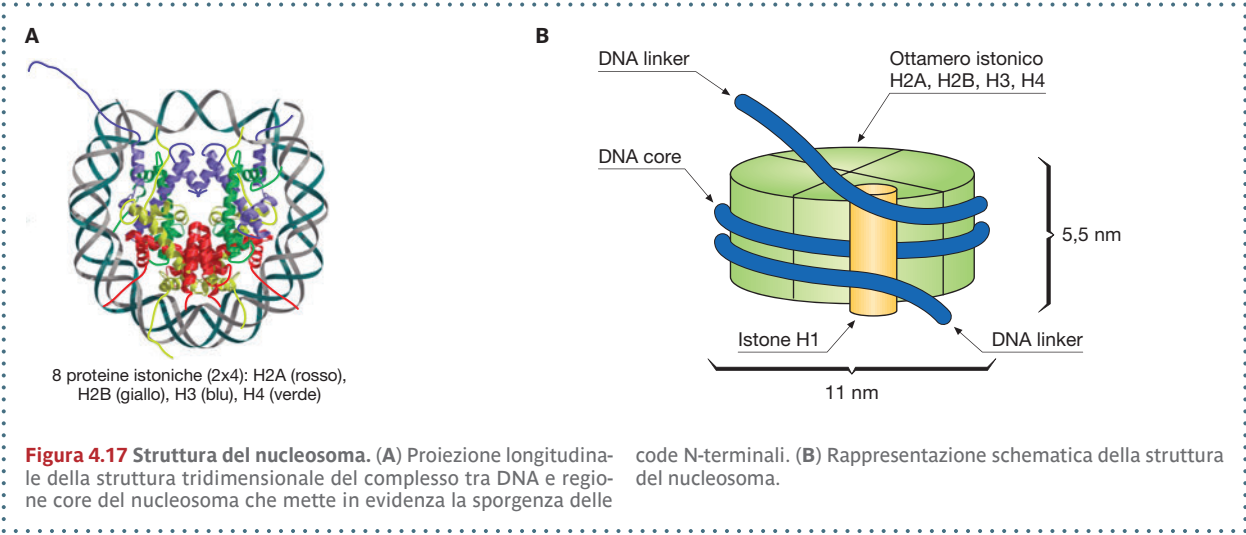


Tabella 4.3 Caratteristiche generali delle proteine istoniche.

Istone	PM (kDa)	Conservazione	Numero di molecole per nucleosoma	Localizzazione
H1	24	+	1	Regione internucleosomica
H2A	14	++	2	Core del nucleosoma
H2B	14	++	2	
H3	15	+++	2	
H4	11	+++	2	

che H2A e H2B sono molto conservati, ma presentano rispetto a H3 e H4 una maggiore variabilità di sequenza amminoacidica tra specie.

Come schematizzato in **figura 4.17B**, l'ottamero di istoni costituisce la struttura centrale o "nucleo" (core) del nucleosoma, a forma di barile o disco o cilindro appiattito del diametro di circa 10 nm (10 nm = 100 Å; ricordiamo che il diametro della doppia elica del DNA è di 20 Å). All'intorno di questo cilindro la doppia elica di DNA compie quasi due giri. A questo nucleo del nucleosoma, ma all'esterno di esso, si trova associata una molecola dell'istone H1, che quindi è presente in quantità pari alla metà di quella degli altri istoni. H1 è, di tutti gli istoni, quello evolutivamente meno conservato, mostrando significative variazioni tra tessuti e tra specie. Circa 140-150 pb di DNA (a seconda della specie) sono strettamente associate al nucleosoma, mentre altre circa 50 pb separano un nucleosoma dall'altro cosicché la cromatina, dipanata e rilassata al massimo, presenta una struttura a collana di perle. L'istone H1, pur non essendo un componente del nucleo del nucleosoma, contribuisce anch'esso alla struttura della cromatina. Esso interagisce con il tratto di DNA tra due nucleosomi (DNA linker) producendo una maggiore adesione del DNA all'ottamero istonico (ulteriore protezione di 20 pb all'azione della nucleasi), ma lega anche una regione mediana delle 147 pb associate all'ottamero istonico. Il legame di H1 aumenta il compattamento del DNA sul nucleosoma, imponendo una costrizione sugli angoli di entrata e uscita del DNA dal nucleosoma.

Le proteine istoniche, che interagiscono con il DNA carico negativamente, sono fortemente basiche in quanto costituite per oltre il 20% dagli amminoacidi carichi positivamente arginina e lisina. Esse presentano una struttura conservata costituita da tre regioni ad α -elica denominata **histone-fold** (**fig. 4.18A**) e da una

regione (coda) N-terminale. La struttura histone-fold media la formazione delle coppie eterodimeriche di H2A con H2B e di H3 con H4, attraverso un'interazione testa-coda. Due coppie di eterodimero H3-H4 si associano poi a formare un tetramero (fig. 4.18B). L'assemblaggio delle proteine istoniche a formare l'ottamero del nucleosoma è un processo ordinato che inizia con il legame al DNA di un tetramero (H3-H4)₂ sul quale vengono poi reclutati due eterodimeri H2A-H2B, che completano l'assemblaggio del nucleosoma (fig. 4.18C e fig. 4.19).

Le code N-terminali degli istoni che, come mostrato schematicamente nella figura 4.20A, sporgono all'esterno del nucleosoma, non sono necessarie per la formazione e il mantenimento della struttura dell'ottamero istonico. Difatti, la digestione con tripsina, che agisce esclusivamente sulle code N-terminali sporgenti rimuovendole, lascia intatto l'ottamero istonico (fig. 4.20B). Queste code N-terminali sono costituite da 20-35 residui e sono regioni bersaglio di una serie di modificazioni chimiche, principalmente acetilazione e metilazione di lisine e fosforilazione di serine, che cambiano la carica netta della molecola proteica diminuendone la basicità. Queste modificazioni hanno un effetto sull'affinità degli istoni per il DNA e quindi sulla stabilità del nucleosoma e hanno un ruolo nei processi di replicazione e trascrizione del genoma e nel controllo epigenetico dell'espressione genica (vedi oltre).

In condizioni di bassa forza ionica e in assenza dell'istone H1 si osserva la struttura meno condensata della cromatina, la **collana di perle**, chiamata anche "fibra da 10 nm" (fig. 4.15A e C), essendo questo il diametro della struttura che si osserva al microscopio elettronico. Se analizziamo la cromatina in condizioni più fisiologiche, essa appare come una "fibra da 30 nm", chiamata anche **sole-**

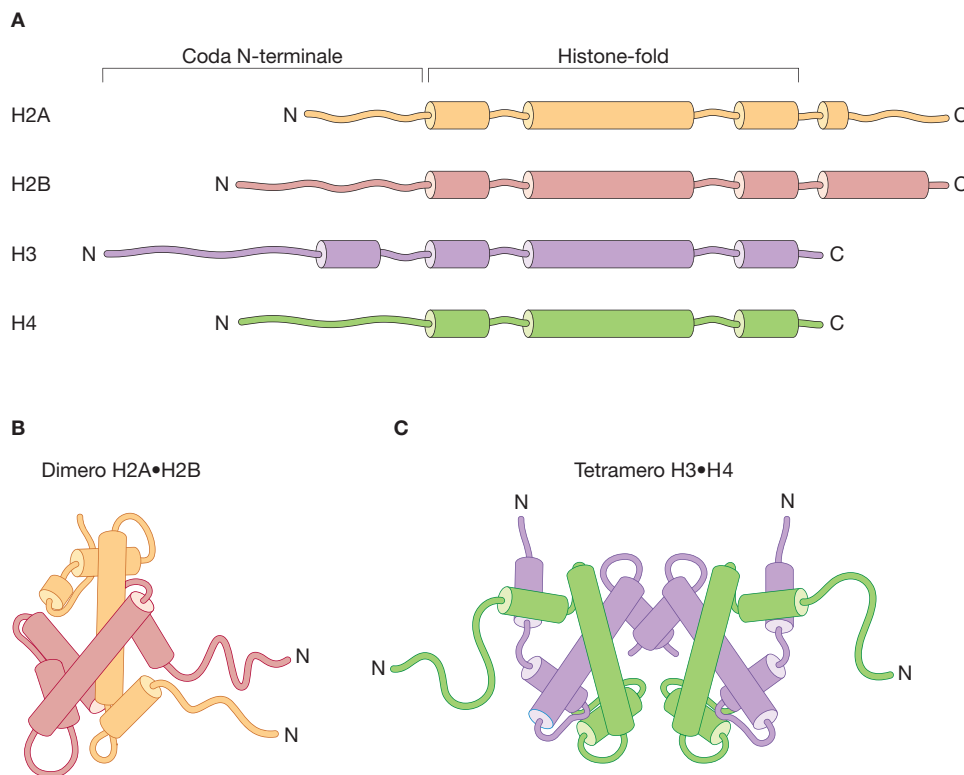
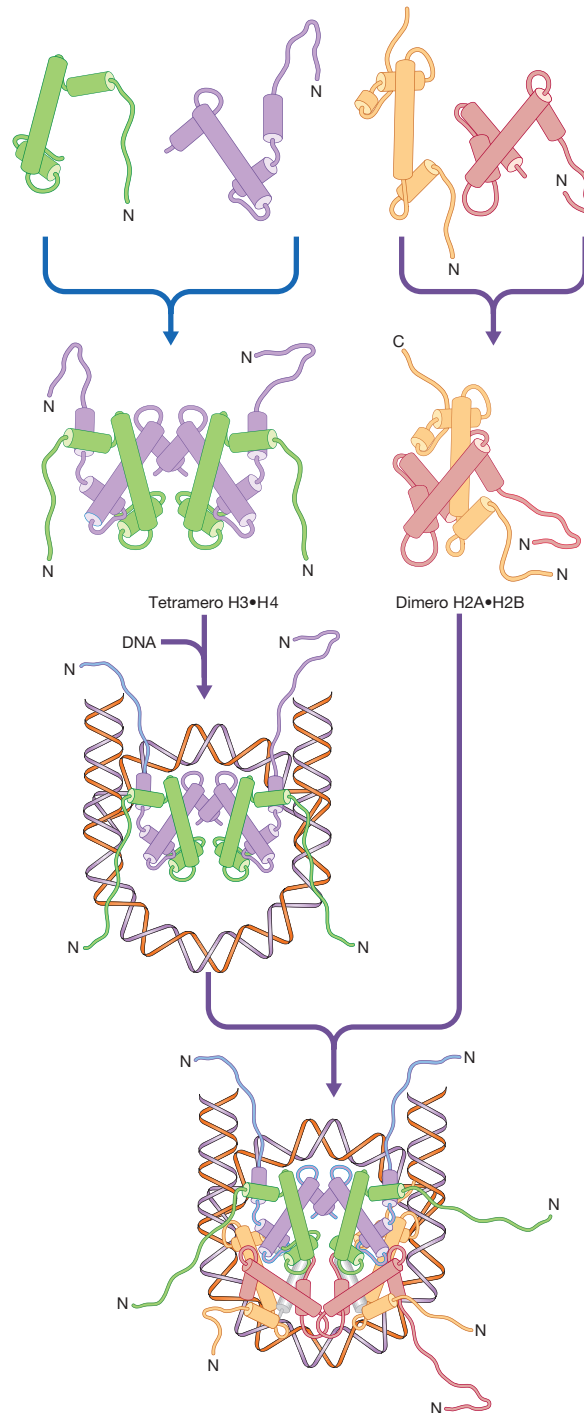


Figura 4.18 Proteine istoniche. In questo tipo di rappresentazione grafica della proteine, le strutture cilindriche rappresentano regioni ad α -elica. (A) I quattro istoni del nucleo del nucleosoma sono costituiti da un dominio conservato, denominato "histone-

fold", che contiene tre α -eliche e una coda N-terminale. (B) Gli istoni H2A e H2B si associano a formare un dimer, mentre le subunità H3 e H4 formano un tetramero (C).

Figura 4.19 La formazione del nucleosoma è un processo ordinato. L'assemblaggio di un nucleosoma inizia con la formazione di tetrameri H3•H4 e di dimeri H2A•H2B. Successivamente un tetramero H3•H4 si lega al DNA formando un complesso su cui vengono poi reclutati due dimeri H2A•H2B.



noide, in cui i nucleosomi appaiono organizzati in una struttura elicoidale con sei nucleosomi per giro (**fig. 4.15B e D**). Le code N-terminali degli istoni contribuiscono alla formazione della fibra da 30 nm mediando interazioni sotto forma di legami idrogeno tra nucleosomi adiacenti. Difatti, la digestione con tripsina, degradando le code N-terminali, impedisce la formazione della fibra da 30 nm.

L'impaccamento della cromatina sui nucleosomi (6 volte; 1 m → 16 cm) e nelle fibre da 30 nm (40 volte; 1 m → 2,5 cm) è ancora molto lontano dall'impac-

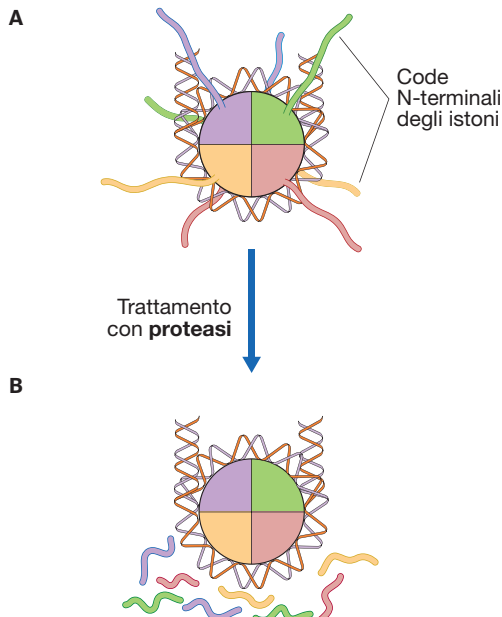
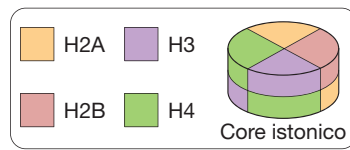


Figura 4.20 Code N-terminali degli istoni. (A) All'esterno del nucleosoma protrudono le code N-terminali degli istoni del nucleo del nucleosoma. (B) La rimozione delle code mediante trattamento con tripsina non altera l'interazione tra DNA e ottamero istonico.

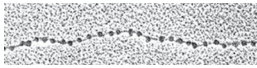
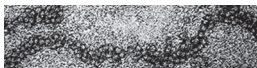
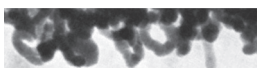
Struttura	Rap. imp.	
Fibra da 10 nm "collana di perle"	6	
Fibra da 30 nm "solenioide"	40	
Anse cromosomiche da 100-400 nm	1000	
Cromosomi mitotici	10000	

Figura 4.21 Compattamento della cromatina. Successivi livelli di organizzazione della cromatina risultano in un crescente rapporto di impacchettamento.

camento necessario (pari a circa 10 000 volte) per contenere il DNA nel nucleo. Il modello strutturale più accreditato realizza questo compattamento attraverso la formazione di anse (30-100 kpb) bloccate alla base da una struttura proteica, la **matrice nucleare** o l'**impalcatura (scaffold) del cromosoma** (vedi par. 4.3). La **topoisomerasi II** e le **proteine SMC** (*Structural Maintenance of Chromosome*) sono componenti essenziali di queste strutture proteiche.

In **figura 4.21** sono illustrati i diversi livelli di organizzazione che corrispondono a rapporti di impacchettamento progressivamente crescenti. L'impacchettamento è circa 1000 volte nelle anse cromosomiche, di cui abbiamo parlato sopra, in cui il solenoide è ulteriormente impacchettato per formare fibre che raggiun-

gono 100–400 nm di diametro. Infine, nei cromosomi mitotici al massimo della contrazione, il rapporto d’impacchettamento raggiunge il valore di circa 10 000.

Posizionamento in fase (phasing) dei nucleosomi

Dopo aver dimostrato che gli istoni possono associarsi a costituire nucleosomi su qualsiasi sequenza di DNA, ci si è chiesti se ci siano delle sequenze su cui essi si assemblano preferenzialmente. Espresso in altri termini, ci si è chiesti se lungo una sequenza di DNA i nucleosomi sono distribuiti in fase, cioè vanno a posizionarsi in punti specifici oppure sono distribuiti a caso. Questo problema è stato affrontato negli anni settanta-ottanta con una serie di studi in cui si analizzavano specifiche regioni genomiche per verificare se i nucleosomi si assemblavano su definite sequenze di DNA. Il problema che ci si poneva, apparentemente semplice e lineare, ha portato a conclusioni abbastanza articolate che possono essere riassunte come segue.

Dal punto di vista del posizionamento dei nucleosomi, non tutte le regioni genomiche sono uguali. In alcune regioni c’è un forte posizionamento in fase, in altre la posizione dei nucleosomi è molto più indeterminata (**fig. 4.22**). Inoltre, per una data regione genica, il posizionamento dei nucleosomi risente anche del particolare stato funzionale, quale la maggiore o minore attività trascrizionale. Per esempio, è stato osservato in *Xenopus* che alcuni geni che codificano per i tRNA sono organizzati in nucleosomi in fase negli eritrociti, dove non sono trascritti, meno in fase nel fegato, dove sono attivamente espressi.

Il posizionamento di un nucleosoma può essere “intrinseco” o “estrinseco”

Si parla di **posizionamento intrinseco** quando, per qualche motivo strutturale, la sequenza del DNA tende a “catturare” un ottamero in una precisa posizione. Il caso più tipico è quello di una sequenza di DNA che presenta una curvatura intrinseca (*vedi* fig. 2.29) dovuta alla particolare sequenza nucleotidica del DNA stesso. Infatti, poiché l’assemblaggio del nucleosoma comporta un avvolgimento del DNA intorno all’ottamero di istoni e poiché questo avvolgimento richiede energia, è chiaro che esso tenderà a posizionarsi preferibilmente nelle regioni del DNA che si presentano intrinsecamente curve.

Si parla invece di **posizionamento estrinseco** quando questo è indotto da qualche altro fattore. Per esempio, la presenza di una proteina legata a una specifica sequenza del DNA impedirà che su questa si assembli un nucleosoma, che sarà quindi spinto ad assemblarsi in posizione adiacente. Ne deriva che diverse proteine legate a posizioni specifiche di un promotore possono “costringere” i nucleosomi ad assemblarsi sulle regioni libere inducendo un loro posizionamento estrinseco. Anche gli stessi nucleosomi, posizionati intrinsecamente o estrinsecamente, possono determinare un posizionamento estrinseco dei nucleosomi adiacenti. Un nucleosoma ben posizionato intrinsecamente induce il posizionamento estrinseco dei nucleosomi adiacenti, ma il posizionamento dei successivi nucleosomi va progressivamente diminuendo mano a mano che ci si allontana dal primo.

Un nucleosoma può assumere diverse posizioni spostandosi lungo la molecola di DNA. Ciò può avvenire con due modalità diverse: **posizionamento traslazionale** e **posizionamento rotazionale**. I termini “traslazione” e “rotazione” si riferiscono al movimento compiuto dal DNA rispetto all’ottamero istonico. Nel caso della traslazione, il contatto ottamero/DNA avviene sempre sulla stessa faccia del DNA. Nel caso della rotazione, il contatto ottamero/DNA si sposta comportando una rotazione del DNA rispetto all’ottamero. Ambedue i tipi di posizionamento, traslazionale e rotazionale, possono avere implicazioni importanti per l’accessibilità del DNA a fattori regolativi. Per esempio, il posizionamento intrinseco di un nucleosoma su una regione di DNA curvo sarà chiaramente

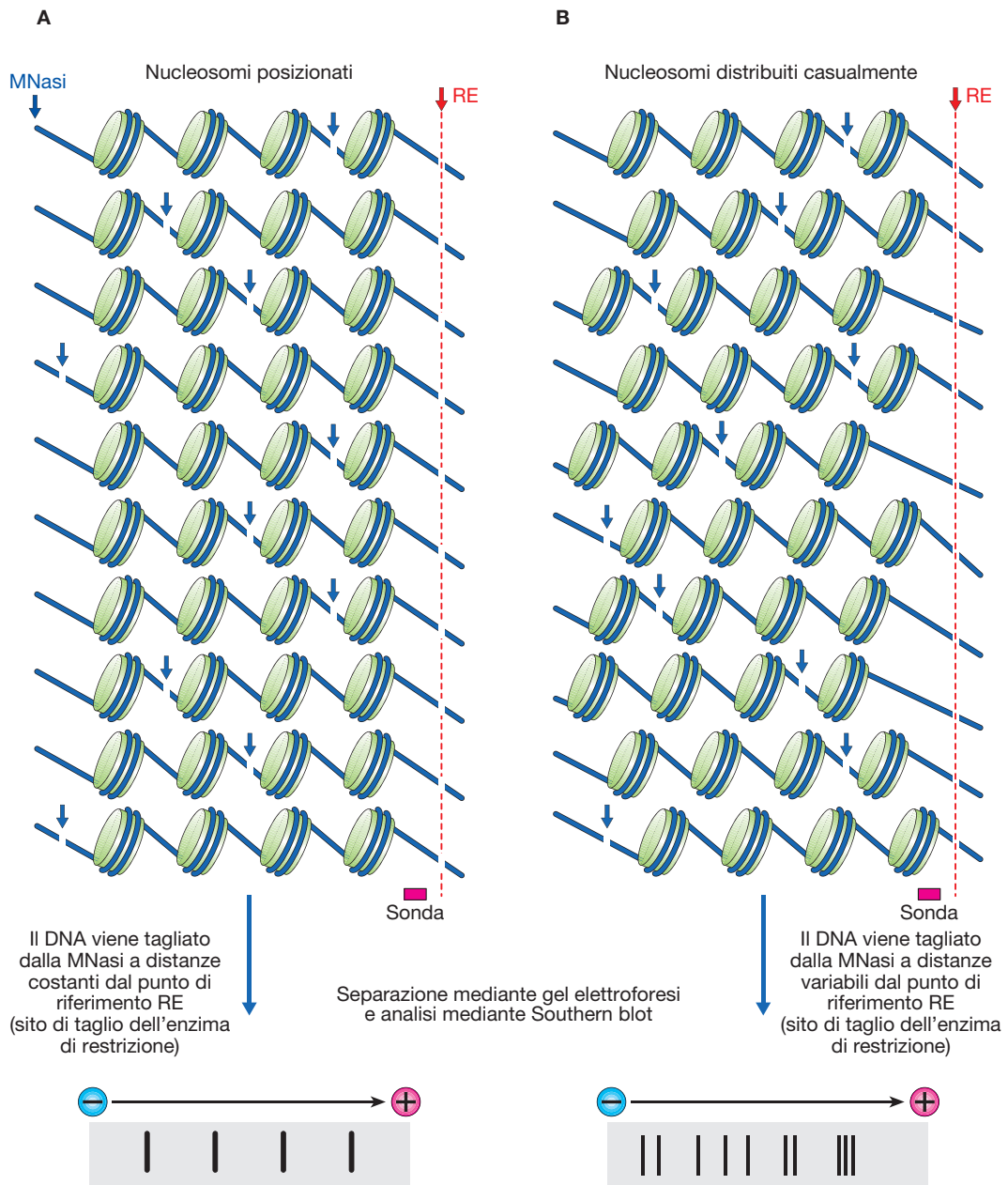


Figura 4.22 Analisi del posizionamento dei nucleosomi nella cellula. I nucleosomi possono essere posizionati in punti specifici (A) o distribuiti in modo casuale (B) lungo la sequenza di DNA. Il posizionamento dei nucleosomi può essere studiato con approcci sperimentali, come quello illustrato in figura. Per determinare i siti di taglio in una determinata regione è necessario scegliere un punto di riferimento, corrispondente al sito di taglio per un enzima di restrizione, in prossimità del quale viene

disegnata una sonda specifica (linea rossa). Il DNA digerito con l'enzima di restrizione (freccia rossa) e la nucleasi micrococcalica MNasi (piccole frecce blu) che taglia in corrispondenza delle regioni linker, viene estratto, separato su gel di agarosio, denaturato e trasferito su una membrana di nitrocellulosa (Southern blot). Trattando la membrana con la sonda specifica marcata (rossa) si osserva l'impronta dei nucleosomi sul DNA, differente nel caso di nucleosomi posizionati in modo casuale o specifico.

di tipo traslazionale garantendo che sempre la stessa faccia della doppia elica del DNA sia esposta a possibili interazioni con, per esempio, fattori di trascrizione. Al contrario, regioni del DNA che non implicano un posizionamento intrinseco del nucleosoma permetteranno la rotazione, rispetto all'ottamero, del DNA, che può così esporre qualsiasi sequenza alle interazioni con fattori regolativi.

Nucleosomi e replicazione del DNA

Cosa succede quando, durante la fase S del ciclo cellulare, il DNA assemblato in nucleosomi viene replicato? Nel capitolo 6 verrà descritta la complessità del processo di replicazione. Una delle complicazioni negli eucarioti è sicuramente dovuta al fatto che il DNA si trova qui organizzato in nucleosomi a formare la fibra cromatinica da 10 nm e questa a sua volta è ulteriormente compattata a formare le strutture di ordine superiore quale la fibra da 30 nm ecc. Gli eventi coinvolti nella replicazione della cromatina sono ancora in gran parte da chiarire, sembra tuttavia molto probabile che la replicazione comporti la disorganizzazione, oltre che delle strutture superiori e della fibra da 30 nm, anche della fibra da 10 nm. La dissociazione dei nucleosomi durante la replicazione è però transitoria. Infatti, i nucleosomi che vengono disassemblati dalla doppia elica parentale vengono rapidamente riassemblati sulle due doppie eliche figlie. L'influenza della struttura della cromatina sulla replicazione è descritta più avanti (*vedi* par. 6.6 e fig. 6.25).

Nucleosomi e trascrizione dei geni

Avendo osservato che gran parte del DNA genomico è organizzato in nucleosomi, ci si è chiesti che relazione ci sia tra questa organizzazione e l'attività genica, o, in altre parole, se i geni sono organizzati in nucleosomi anche quando sono attivamente trascritti.

Sicuramente qualche problema ci deve essere. Infatti, se durante la trascrizione la struttura nucleosomiale fosse mantenuta, l'RNA polimerasi dovrebbe girare intorno ai nucleosomi stessi, quasi due giri per ogni 200 coppie di basi. Questo sembra abbastanza improbabile, anche in considerazione del fatto che l'RNA polimerasi e il nucleosoma sono grandi strutture macromolecolari di dimensioni più o meno equivalenti. Inoltre, anche l'osservazione che durante la trascrizione l'RNA polimerasi interagisce strettamente con circa 50 coppie di basi, 10-15 delle quali vengono aperte per permettere la copiatura dello stampo, suggerisce che l'organizzazione del nucleosoma non possa restare immutata durante la trascrizione.

Le tecniche usate, soprattutto tra la metà degli anni settanta e la metà degli anni novanta, per studiare questo problema sono state principalmente: (1) l'osservazione al microscopio elettronico della cromatina, che permette di distinguere visivamente le regioni trascritte e quelle organizzate in nucleosomi; (2) la digestione della cromatina con enzimi endonucleolitici di vario tipo, seguita dall'analisi elettroforetica dei frammenti di DNA prodotti, che permette di individuare le regioni di maggiore e minore sensibilità alle nucleasi e quindi di risalire alla minore o maggiore compattezza dell'organizzazione cromatinica. I risultati ottenuti in tanti anni di studio sono abbastanza contraddittori e difficilmente schematizzabili.

La **figura 4.23** mostra due fotografie al microscopio elettronico di due situazioni estreme: (A) i geni che nel nucleolo codificano per l'rRNA ("alberi di Natale" di Miller, *vedi* cap. 12), trascritti molto attivamente; (B) il cromosoma circolare del virus SV40, che è trascritto molto meno attivamente rispetto ai geni per gli rRNA. Si può osservare che il cromosoma di SV40, che in questo caso presenta una sola molecola di RNA in corso di trascrizione, è per il resto organizzato in nucleosomi; al contrario, i geni per l'rRNA che presentano numerose molecole di RNA in corso di trascrizione, una adiacente all'altra, non presentano un'evidente organizzazione nucleosomica.

Gli studi fatti digerendo la cromatina con enzimi endonucleolitici quali la nucleasi micrococcica, la DNasi I e la DNasi II, hanno portato alle seguenti conclusioni:

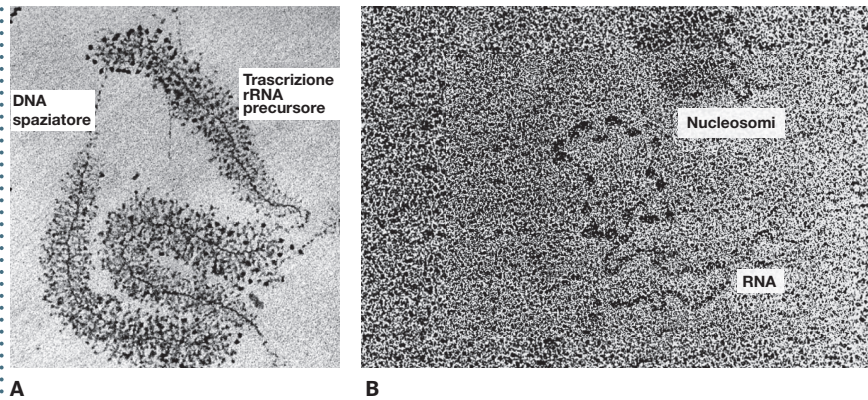


Figura 4.23 Nucleosomi e trascrizione. Immagini al microscopio elettronico del processo trascrizionale dei geni per i precursori dell'rRNA (A) e del minicromosoma del virus SV40 (B).

1. in genere i geni che vengono trascritti contengono nucleosomi come le sequenze non trascritte;
2. i nucleosomi che in genere sono posizionati in fase su un gene non trascritto, perdono il posizionamento quando il gene viene attivato trascrizionalmente;
3. i geni trascrizionalmente attivi vengono digeriti dalle nucleasi più rapidamente delle regioni non trascritte, suggerendo che le regioni cromosomiche trascrizionalmente attive, pur mantenendo l'organizzazione nucleosomica, siano in qualche modo più dipanate e quindi più esposte alla digestione enzimatica.

Questi e altri studi simili hanno suggerito la conclusione, limitata ma importante, che le regioni di DNA trascritte possano essere organizzate in nucleosomi come quelle non trascritte e che, quindi, non esistano organizzazioni strutturali alternative. L'RNA polimerasi, avanzando sul DNA durante il processo di trascrizione, scalza i nucleosomi, che poi si riassemblano sul DNA subito dietro la polimerasi, dopo il suo passaggio. Durante questo processo, mentre può essere necessario che l'istone H1 venga rimosso, l'ottamero istonico che costituisce il nucleo del nucleosoma non viene dissociato ma si mantiene come struttura unitaria pronto a rilegare il DNA dopo che la polimerasi è transitata. Solo quando la trascrizione è molto attiva, come è il caso dei geni per l'rRNA, i nucleosomi non hanno il tempo di riassemblarsi tra un evento di trascrizione e il successivo.

Lo studio della sensibilità alle DNasi delle varie regioni cromosomiche nella cromatina ha permesso di identificare due fenomeni: una variazione generale della sensibilità alle DNasi di una regione cromosomica quando questa viene trascrizionalmente attivata e la presenza di siti "ipersensibili" alle DNasi in particolari regioni coinvolte nella regolazione genica.

1. **Sensibilità alle nucleasi.** L'attivazione trascrizionale di un gene comporta una alterazione della sua organizzazione cromatinica (a prescindere dalla disorganizzazione della struttura nucleosomica che può essere momentaneamente causata dal passaggio dell'RNA polimerasi). Questi cambiamenti strutturali della cromatina trascritta si rivelano come un'aumentata sensibilità dell'intera regione genica alla degradazione ad opera della DNasi I.
2. **Siti ipersensibili.** Oltre ai sopra descritti cambiamenti generali che avvengono con l'attivazione trascrizionale dei geni, si possono avere anche dei cambiamenti strutturali in siti specifici, associati per lo più con l'inizio della trascrizione. Questi sono stati scoperti facendo delle digestioni di preparazioni di cromatina con endonucleasi molto diluite. I primissimi tagli prodotti sul DNA cromosomico dalle nucleasi individuano dei siti che sono particolarmente esposti non essendo coinvolti nell'interazione con i nucleosomi. Questi

siti, chiamati “siti ipersensibili”, sono in genere correlati all’espressione genica e si trovano per lo più nelle regioni regolative a monte dei geni, dove l’interazione del DNA con fattori di trascrizione generali o regolativi impedisce la normale organizzazione nucleosomica.

In anni più recenti nuove tecniche e nuovi modelli sperimentali hanno permesso di studiare in maniera più puntuale le relazioni tra regolazione della trascrizione e organizzazione cromatinica e nucleosomica, argomento che però rimandiamo ai paragrafi 11.7 e 17.1.

Modificazioni post-traduzionali degli istoni e modulazione dell’attività trascrizionale della cromatina

Come abbiamo già visto, le code N-terminali degli istoni non contribuiscono alla struttura e alla stabilità del nucleosoma, tuttavia hanno un ruolo fondamentale nella modulazione della struttura e del grado di accessibilità della cromatina. Esse, infatti, possono subire una serie di modificazioni post-traduzionali, mediate da specifiche attività enzimatiche, che comprendono l’acetilazione e la metilazione delle lisine (K) e delle arginine (R), la fosforilazione di serine (S) e treonine (T), l’ubiquitinazione e la sumoilazione delle lisine, come pure la ADP-ribosilazione (fig. 4.24). A incrementare il livello di complessità di queste modificazioni si consideri che ciascun residuo di lisina può aggiungere uno, due o tre gruppi metilici, mentre l’arginina può essere mono- o dimetilata.

Dato che la cromatina è lo stampo fisiologico di tutti i processi del DNA eucariotico, è evidente che differenti modificazioni degli istoni portano a diverse conseguenze funzionali. Per esempio, la combinazione di acetilazione su K8 di H4 (cioè della lisina in posizione 8 dell’istone H4) e K14 di H3 e la fosforilazione su S10 di H3 sono associate con uno stato trascrizionalmente attivo. Al contrario, la trimetilazione su K9 di H3 e l’assenza di acetilazione su H3 e H4 sono generalmente associate con la repressione della trascrizione. Specifici pattern di modificazione degli istoni sono anche associati a dinamiche più generali dello stato della cromatina. Per esempio, la deacetilazione dell’istone H4 in K4 e K12



Figura 4.24 Modificazioni post-traduzionali dei quattro istoni dell’ottamero nucleosomico. Le code N-terminali degli istoni

possono essere metilate (Me) su lisine e arginine, fosforilate (P) su serine e treonine, ubiquitate (Ub) e acetilate (Ac) su lisine.

è associata con la deposizione degli istoni in fase S, mentre la fosforilazione degli istoni H2A (S1 e T119) e H2B (T3, S10 e S28) caratterizza la cromatina mitotica condensata. Queste, e molte altre osservazioni, hanno suggerito l'esistenza di un **codice istonico**, per cui a un determinato assortimento di modificazioni post-traduzionali dovrebbe corrispondere la stessa risposta funzionale. Specifici assortimenti di modificazioni istoniche possono attivare (o reprimere) la trascrizione reclutando complessi di modificazione o rimodellamento della cromatina (*vedi* parr. 11.7 e 17.2 per ulteriori dettagli).

Le modificazioni chimiche degli istoni sono prodotte da grandi famiglie di complessi proteici con attività enzimatica che comprendono acetil transferasi (HAT), deacetilasi (HDAC), metil transferasi (HMT) e demetilasi, e chinasi o fosfatasi, ciascuna specializzata per specifici residui amminoacidici sugli istoni (*vedi* anche cap. 19). Tali attività sono anche tra loro strettamente interconnesse, dato che uno specifico tipo di marcatura istonica può indurne o reprimere altre. Vi sono, infatti, specifici moduli proteici che riconoscono particolari tipi di modificazioni: per esempio, i **cromodomini** riconoscono le lisine metilate, mentre i **bromodomini** si legano in modo specifico a lisine acetilate (*vedi* anche par. 19.5). Anche i complessi deputati al rimodellamento della cromatina contengono questo tipo di domini funzionali.

Recentemente, alcuni studi hanno dimostrato che anche alcuni piccoli RNA non codificanti possono modulare le modificazioni degli istoni e quindi lo stato della cromatina.

Lo stato della cromatina può essere modificato anche dall'incorporazione di varianti non canoniche delle proteine istoniche. Per esempio, H2A.Z è una variante di H2A che è generalmente presente nei nucleosomi della cromatina attivamente trascritta, in quanto impedisce la formazione di regioni eterocromatiche. Una seconda variante di H2A, H2A.X, nello stato fosforilato, è un marcatore della presenza di lesioni al DNA. Sono note anche varianti dell'istone H3 che hanno acquisito nel corso dell'evoluzione delle funzioni specializzate. CENP-A è una variante di H3 associata a nucleosomi localizzati in regioni centromeriche: avendo una coda N-terminale più estesa, CENP-A permette una più efficace interazione con le proteine del cinetocore, mediando la segregazione cromosomica. H3.3, un'altra variante di H3, è invece associata a regioni attivamente trascritte e come H2A.Z previene la formazione di tratti di eterocromatina.

I cambiamenti strutturali della cromatina che modulano l'espressione dei geni sono trattati anche nel paragrafo 17.2. Va infine ricordato che lo stato della cromatina è anche correlato al profilo di metilazione del DNA, che viene trattato nel paragrafo 17.1.

Lecture di approfondimento

- Oganesian, L. e Bryan, M.T. (2007), «Physiological relevance of telomeric G-quadruplex formation: a potential drug target», *Bioessays*, **29**, pp. 155-165.
- Peterson, C.L. e Laniel, M.A. (2004), «Histones and histone modifications», *Current Biology*, **14**(14), pp. 546-551.
- Schueler, M.G. e Sullivan, B.A. (2006), «Structural and functional dynamics of human centromeric chromatin», *Annual Review of Genomics and Human Genetics*, **7**, pp. 301-313.

5

Genomi procariotici ed eucariotici

5.1 Era genomica

5.2 Genomi procariotici

5.3 Genomi eucariotici

5.4 Genoma mitocondriale

5.5 Genoma dei cloroplasti

5.6 Genoma dei virus

5.1 Era genomica

Lo studio della struttura e organizzazione dei genomi ha fatto in questi decenni due salti qualitativi stimolati dai progressi metodologici di analisi dei genomi. Infatti, negli anni sessanta si poteva solo analizzare il DNA totale preparato da cellule o tessuti ottenendo informazioni molto generiche come, per esempio, la composizione in basi media, la presenza di sequenze ripetute e altamente ripetute ecc. Un importante salto qualitativo è stato possibile, a partire dal 1975, con la possibilità di clonare (isolare) frammenti cromosomici ragionevolmente piccoli e analizzarne la sequenza nucleotidica (*vedi* par. 20.7). Dobbiamo a questo periodo lo studio approfondito della struttura di singoli geni, dell'organizzazione delle famiglie geniche, delle sequenze trasponibili e anche la scoperta degli introni. Successivamente, l'avvento dell'“era genomica”, iniziata nel 1995 con la pubblicazione sulla prestigiosa rivista *Science* del genoma di *Haemophilus influenzae*, ha avviato una profonda rivoluzione nella ricerca biologica aprendo prospettive di indagine prima nemmeno immaginabili. Gli studi biomolecolari hanno così assunto una dimensione “omica”, dove questo termine, utilizzato come suffisso in molti ambiti, sta a rappresentare la dimensione di scala dell'indagine, relativa all'intero complesso delle unità elementari di un dato sistema (ad es. genoma, trascrittoma, proteoma, interattoma ecc.).

Dall'inizio dell'era genomica il costante e progressivo sviluppo delle tecnologie di indagine sperimentale, particolarmente nel campo delle tecniche di sequenziamento del DNA, ha reso possibile il sequenziamento completo del genoma di moltissimi organismi procariotici ed eucariotici, compreso il genoma umano.

Il recentissimo avvento delle piattaforme di sequenziamento di nuova generazione (*vedi* par. 20.8) produrrà un'ulteriore significativa accelerazione delle capacità di sequenziamento con una concomitante costante riduzione dei costi.

La disponibilità di molti genomi, consentendone l'analisi comparativa, rende molto più agevole l'interpretazione delle informazioni in essi contenute. Gli studi di genomica comparata ci permettono di osservare gli effetti dell'evoluzione. Ciò che è conservato, ovvero preservato nel corso dell'evoluzione dalla selezione naturale, è molto probabile che abbia una funzione biologica fondamentale, comune a più organismi.

L'informazione presente nei genomi, come pure i prodotti e i processi biologici che risultano dalla loro espressione, possono essere interpretati molto più agevolmente alla luce del processo evolutivo. Infatti, come affermava Theodosius Dobzhansky, un importante genetista russo, “nulla in Biologia ha senso se non alla luce dell'evoluzione”.

5.2 Genomi procariotici

Caratteristiche, struttura e plasticità dei genomi procariotici

I procarioti sono organismi unicellulari microscopici che vivono in una grande varietà di ambienti (suolo, acque, altri organismi). Anche se oggi sono note circa 4000 specie, si stima che questo numero sia compreso in realtà tra 400 000 e 4 000 000.

La definizione di “specie” nel caso dei batteri è piuttosto arbitraria e normalmente si basa su una serie di caratteri morfologici, biochimici e molecolari, come per esempio quella basata sull'analisi filogenetica dell'rRNA 16S.

La classificazione molecolare suddivide i procarioti in due domini: **Batteri** (o eubatteri) e **Archeobatteri**, che con gli **Eucarioti** formano i tre rami principali dell'Albero delle Vita (*Tree of Life*) (fig. 5.1).

I procarioti sono quindi suddivisi in due grandi raggruppamenti: gli archeobatteri e gli eubatteri. Gli archeobatteri, che normalmente risiedono in habitat estremi, presentano una serie di caratteristiche e proprietà che li differenziano sostanzialmente dagli eubatteri:

- le proteine “informazionali” (cioè quelle implicate in replicazione, trascrizione e traduzione) sono più simili alle controparti eucariotiche che a quelle eubatteriche (ad es. l'RNA polimerasi è simile nella struttura all'RNA polimerasi II eucariotica);
- posseggono proteine simili a istoni per l'impacchettamento del DNA;

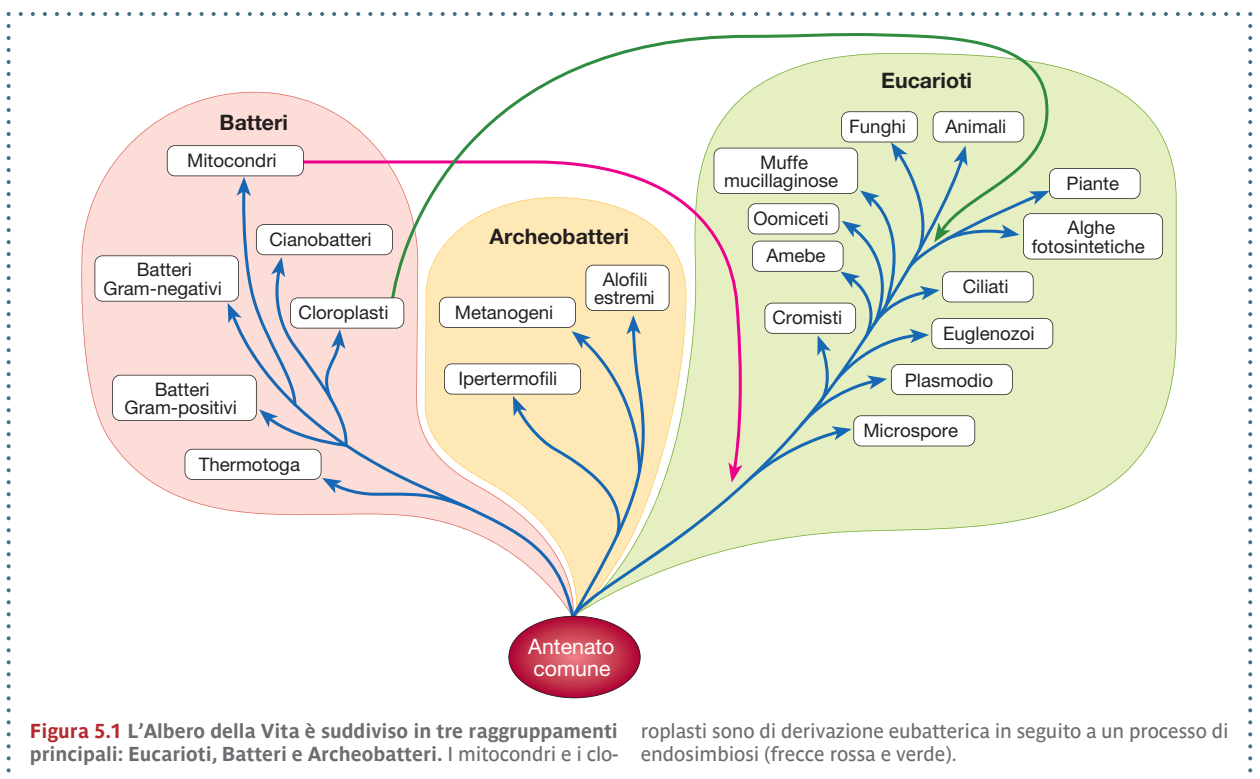


Figura 5.1 L'Albero della Vita è suddiviso in tre raggruppamenti principali: Eucarioti, Batteri e Archeobatteri. I mitocondri e i cloroplasti sono di derivazione eubatterica in seguito a un processo di endosimbiosi (freccia rossa e verde).

- la struttura del promotore è simile a quella dei promotori eucariotici, con la TATA box in posizione –30 rispetto al sito di inizio della trascrizione, mentre negli eubatteri la TATA box – detta anche Pribnow box – si trova in posizione –10 (*vedi* cap. 9);
- la traduzione inizia con una metionina (mentre inizia con formil-metionina negli eubatteri);
- alcuni geni per tRNA e rRNA possono contenere introni.

I genomi procariotici sono generalmente costituiti da un singolo cromosoma di forma circolare. In molte specie sono presenti anche piccoli DNA circolari extra-cromosomici contenenti geni addizionali.

Vi sono anche alcuni casi di genomi di forma lineare. Per esempio, il genoma della spirocheta *Borrelia burgdorferi*, il batterio responsabile della malattia di Lyme, ha un genoma lineare di 0,9 Mpb che presenta telomeri costituiti da forcine chiuse covalentemente e da 21 plasmidi lineari e circolari (complessivamente 0,6 Mpb).

I genomi batterici mostrano una notevole variabilità delle dimensioni, normalmente correlata al loro stile di vita (**fig. 5.2**). Si va da 144 kpb nel batterio endosimbionte *Candidatus Hodgkinia cicadicola* a 13 000 kpb in *Sorangium cellulosum*, un mixobatterio che vive nel suolo in grado di produrre epotilone, un farmaco utilizzato nella cura dei tumori (**tab. 5.1**). In particolare, i batteri simbiotici intracellulari obbligati posseggono genomi piccoli molto stabili, che raramente presentano riarrangiamenti genomici.

I genomi archeobatterici hanno dimensioni poco variabili, generalmente comprese tra 1,5 e 2,5 Mpb.

I genomi procariotici mostrano una “plasticità” molto elevata, per cui i genomi di individui appartenenti a una stessa specie possono avere dimensioni significativamente differenti. La componente genomica comune è denominata **endogenoma**, mentre la frazione variabile è estremamente diversa da ceppo a ceppo.

Per esempio, i circa 20 genomi sinora sequenziati di *Escherichia coli* hanno dimensioni comprese tra 4,6 e 5,5 Mpb. In particolare, il ceppo patogeno O157 ha 1,34 Mpb specifiche per fattori di virulenza, assenti nel ceppo di laboratorio K12, che a sua volta contiene 0,53 Mpb assenti nel ceppo O157.

L'elevata plasticità dei genomi batterici è correlata alla possibilità di eventi di trasferimento genico laterale, nei quali del materiale genetico viene trasferito da un individuo a un altro, molto più frequenti nei batteri rispetto agli eucarioti.

Figura 5.2 Distribuzione delle dimensioni dei genomi di 526 specie di batteri. [Fonte: Comprehensive Microbial Resource, http://cmr.jcvi.org/cgi-bin/CMR/shared/Genomes.cgi?bacteria_only=1, ottobre 2009.]

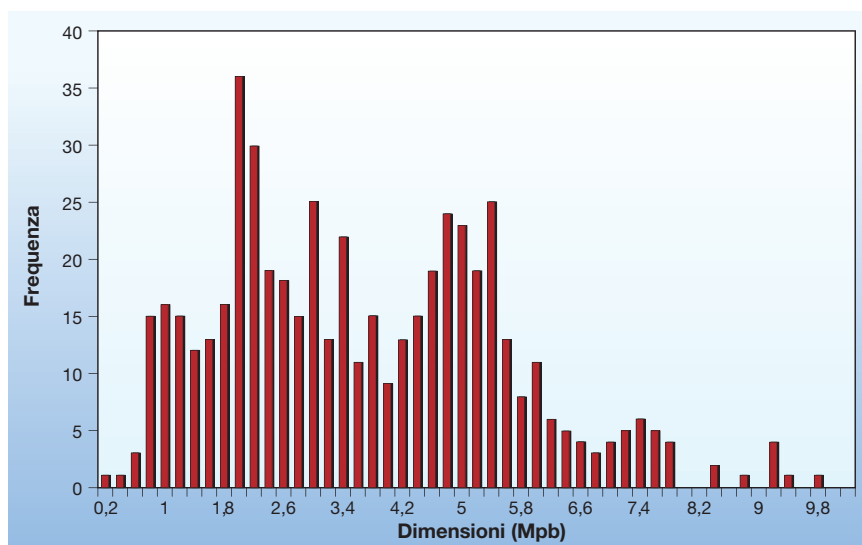


Tabella 5.1 Dimensioni e numero di geni di alcuni genomi procariotici. È anche riportato il numero di cromosomi, la percentuale di G+C e l'identificativo (Accession) della banca dati NCBI/EMBL.

Specie	Cromosomi	Dimensione genoma (nt)	Numero di geni	% GC	Accession
<i>Candidatus Hodgkinia cicadicola</i>	1	143 795	169	58,4	CP001226
<i>Mycoplasma genitalium</i>	1	580 076	476	31,7	L43967
<i>Rickettsia prowazekii</i>	1	1 111 523	834	29,0	AJ235269
<i>Borrelia burgdorferi</i>	22	1 519 856	1639	28,2	AE000783-794; AE0001575-1584
<i>Helicobacter pylori</i>	1	1 576 758	1528	39,2	FM991728
<i>Haemophilus influenzae</i>	1	1 813 033	1623	38	CP000671
<i>Streptococcus pneumoniae</i>	1	2 111 882	2073	39,7	CP000920
<i>Neisseria meningitidis</i>	1	2 153 416	2020	51,7	CP000381
<i>Staphylococcus aureus</i>	1	2 820 462	2632	32,8	BA000033
<i>Escherichia coli</i> str. K12 DH10B	1	4 686 137	4200	50,8	CP000948
<i>Clostridium botulinum</i>	1	3 863 450	3552	28,2	CP000726
<i>Sorangium cellulosum</i>	1	13 033 779	9384	71,4	AM746676

Lo scambio di materiale genetico può avvenire attraverso tre meccanismi principali: (1) trasformazione; (2) coniugazione e (3) trasduzione. Nella trasformazione il batterio ricevente può acquisire DNA esogeno nella condizione in cui sia reso “competente” per questo trasferimento (vedi anche par. 20.4). Nella coniugazione il trasferimento avviene attraverso il contatto fisico cellula-cellula, che vede il concorso di plasmidi coniugativi come il plasmide F. Infine, nella trasduzione il trasferimento di materiale genetico è mediato da un'infezione fagica. Gli eventi di trasferimento genico laterale possono conferire al batterio ricevente particolari proprietà, come la resistenza ad antibiotici, o specifiche proprietà di virulenza. In questo caso il tratto genico che conferisce le proprietà patogene, che normalmente mostra un contenuto di G+C atipico rispetto al resto del genoma, viene denominato **isola di patogenicità**.

Contenuto genico e composizione in basi dei genomi procariotici

I genomi dei procarioti hanno una densità genica molto elevata. In media i geni codificanti, che esprimono proteine della lunghezza media di 300 aa, occupano l'85% del genoma. Inoltre, i geni procariotici non sono interrotti da introni e sono generalmente organizzati in unità trascrizionali policistroniche (ad es. operone *lac*). In questo modo uno stesso promotore può controllare l'espressione di più geni tra loro funzionalmente correlati contribuendo così a rendere il genoma più compatto, riducendo al minimo l'estensione delle regioni regolatorie.

Il numero di geni, così come le dimensioni del genoma, riflettono lo stile di vita del batterio. I parassiti specializzati hanno un numero di geni ridotto (ad es. il genoma di *Mycoplasma genitalium* codifica per circa 500 geni), mentre i batteri generalisti hanno un numero di geni molto maggiore, tipicamente compreso tra 4000 e 5000.

Il numero dei geni negli archeobatteri è meno variabile e generalmente compreso tra 1700 e 2900.

L'elevata plasticità dei genomi procariotici si riflette nel fatto che l'ordine dei geni lungo il genoma è poco conservato tra specie e raggruppamenti tassonomici differenti. Pertanto, gruppi di geni contenuti nello stesso operone nel genoma di una specie possono essere dispersi nel genoma di un'altra specie.

La composizione in basi, misurata come G+C%, dei genomi procariotici è molto variabile e compresa tra il 25 e il 75% (fig. 5.3).

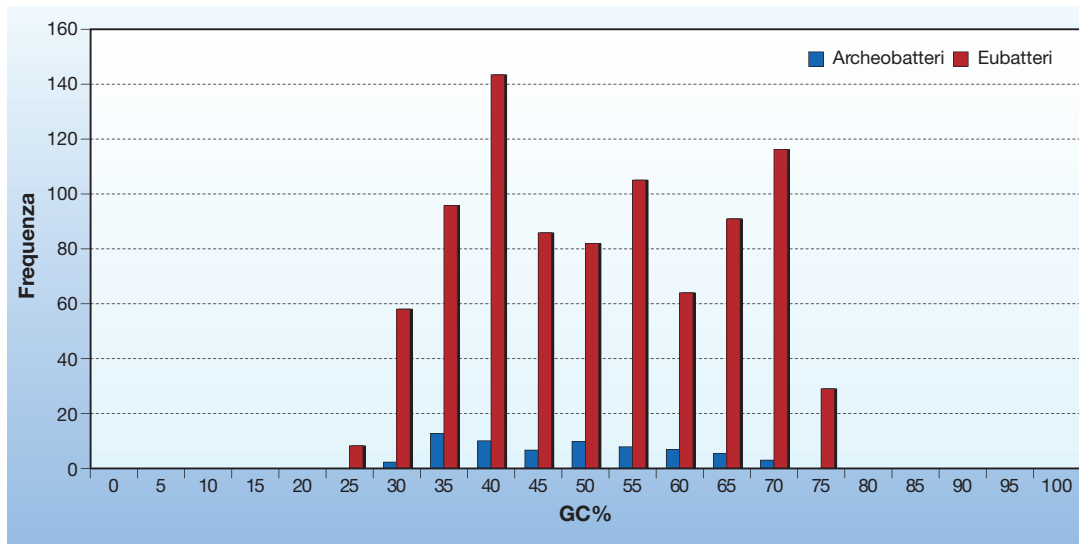


Figura 5.3 Distribuzione della composizione in GC% dei genomi di 878 specie di eubatteri e 66 archeobatteri. (Fonte: NCBI Ge-

nomes, <http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>, ottobre 2009.)

Un'altra proprietà composizionale misurabile dei genomi è l'asimmetria, che misura la distribuzione delle basi complementari (G/C o A/T) su uno stesso filamento. L'**asimmetria composizionale** (o *skew*) si può facilmente misurare come:

$$AT_{\text{skew}} = \frac{A - T}{A + T} \qquad GC_{\text{skew}} = \frac{G - C}{G + C}$$

e assume valori compresi tra -1 e $+1$. I genomi batterici sono globalmente marcatamente simmetrici, con valori di AT_{skew} e $GC_{\text{skew}} < 0,01$. È possibile, tuttavia, osservare specifici tratti del genoma che mostrano un significativo livello di asimmetria. In particolare, nei genomi batterici con una singola origine di replicazione si osserva una transizione del segno dell'indice di asimmetria proprio in corrispondenza dell'origine della replicazione. Questo è probabilmente dovuto a una differente efficienza dei sistemi di riparazione tra *leading* e *lagging strand* (vedi cap. 6), che si invertono proprio in corrispondenza dell'origine di replicazione.

Le proprietà composizionali del genoma influenzano anche la strategia di utilizzo dei codoni sinonimi. I genomi con un elevato contenuto di G+C (A+T) tendono a preferire codoni sinonimi con G o C (A o T) in terza posizione (**fig. 5.4**). La strategia di uso dei codoni, oltre che dalle proprietà composizionali del genoma, può essere influenzata dalla disponibilità di specifici tRNA. Si osserva, infatti, in alcuni batteri come *E. coli*, una più marcata asimmetria nell'uso dei codoni sinonimi per i geni con un elevato livello di espressione. Per rendere ottimale l'efficienza della traduzione vengono difatti utilizzati quasi esclusivamente i codoni corrispondenti ai tRNA isoaccettori più abbondanti.

Geni non codificanti proteine ed elementi mobili nei genomi procariotici

La porzione non codificante proteine dei genomi procariotici, che come abbiamo già visto corrisponde a circa il 15% del genoma, oltre a contenere i geni per rRNA e tRNA, contiene geni per altri RNA con funzione regolatoria (**sncRNA**, *Small Non Coding RNA*) che hanno una funzione simile ai piccoli microRNA eucariotici.

		Seconda posizione																
		U				C				A						G		
Prima posizione (estremità 5')	U	UUU	Phe	4,88%	8702	UCU	Ser	3,08%	5491	UAU	Tyr	3,06%	5485	UGU	Cys	0,99%	1780	U
		UUC	Phe	0,42%	759	UCC	Ser	0,20%	357	UAC	Tyr	0,46%	821	UGC	Cys	0,21%	382	C
		UUA	Leu	6,64%	11903	UCA	Ser	1,86%	3328	UAA	STOP	0,24%	431	UGA	STOP	0,03%	59	A
		UUG	Leu	0,92%	1650	UCG	Ser	0,22%	402	UAG	STOP	0,03%	55	UGG	Tro	0,89%	1594	G
	C	CUU	Leu	1,22%	2185	UCU	Pro	1,44%	2573	CAU	His	1,71%	3059	CGU	Arg	1,18%	2110	U
		CUC	Leu	0,12%	226	UCC	Pro	0,15%	280	CAC	His	0,22%	404	CGC	Arg	0,15%	285	C
		CUA	Leu	0,80%	1435	UCA	Pro	1,18%	2107	CAA	Gin	2,69%	4808	CGA	Arg	0,56%	1000	A
		CUG	Leu	0,15%	284	UCG	Pro	0,16%	297	CAG	Gin	0,31%	553	CGG	Arg	0,04%	83	G
	A	AUU	Ile	6,53%	11651	ACU	Thr	1,99%	3562	AAU	Asn	6,23%	11121	AGU	Ser	1,50%	2677	U
		AUC	Ile	0,78%	1399	ACC	Thr	0,21%	392	AAC	Asn	1,03%	1838	AGC	Ser	0,24%	444	C
		AUA	Ile	4,35%	7771	ACA	Thr	1,96%	3495	AAA	Lys	9,74%	17369	AGA	Arg	1,57%	2799	A
		AUG	Met	2,06%	3688	ACG	Thr	0,21%	387	AAG	Lys	0,85%	1518	AGG	Arg	0,12%	214	G
	G	GUU	Val	2,22%	3970	GCU	Ala	2,00%	3579	GAU	Asp	3,74%	6668	GGU	Gly	2,29%	4090	U
		GUC	Val	0,28%	503	GCC	Ala	0,20%	362	GAC	Asp	0,46%	832	GGC	Gly	0,27%	489	C
		GUA	Val	1,90%	3403	GCA	Ala	1,92%	3428	GAA	Glu	5,13%	9163	GGA	Gly	2,52%	4500	A
		GUG	Val	0,30%	544	GCG	Ala	0,20%	368	GAG	Glu	0,42%	766	GGG	Gly	0,23%	416	G

545 ORF totali, 178 274 codoni

Codoni [% Frequenza codoni] [Numero di codoni]

545 ORF totali, 178 274 codoni

Codoni [% Frequenza codoni] [Numero di codoni]

Figura 5.4 Strategia di utilizzo dei codoni nel batterio *Buchnera aphidicola* Sg. Il genoma di questo batterio ha un basso contenuto in G+C% (25%) e si può osservare come vengano preferiti i codoni

sinonimi con A o T in terza posizione. [Fonte: Comprehensive Microbial Resource, <http://cmr.jcvi.org/>, ottobre 2009.]

Gli snRNA vengono prodotti a partire da promotori canonici e mostrano normalmente una elevata propensione a formare strutture secondarie complesse. Senza subire alcun processamento si legano alla chaperonina Hfq in forma esamerica e formano un appaiamento complementare con l'mRNA bersaglio. Questa interazione può avere diversi effetti sull'espressione del trascritto, quali l'attivazione o la repressione della traduzione, la stabilizzazione o la degradazione dell'mRNA.

I genomi batterici contengono, come vedremo anche per gli eucarioti, elementi mobili (*Mobile Genetic Element*, MGE), o trasposoni, in grado di mediare il trasferimento di materiale genetico all'interno dello stesso genoma o tra genomi differenti. Come descriveremo nel capitolo 8 (vedi par. 8.3), gli elementi mobili più semplici sono denominati elementi IS (*Insertion Sequence*), hanno dimensioni variabili da 600 a oltre 3000 pb e sono costituiti da brevi sequenze ripetute e invertite che fiancheggiano un gene codificante per una trasposasi che media la loro stessa trasposizione. I genomi procariotici possono anche contenere elementi mobili più complessi come i trasposoni composti, fiancheggiati da due elementi IS, e che possono contenere vari geni che possono conferire all'ospite resistenza ad antibiotici o esprimere fattori di virulenza (vedi par. 8.3 e fig. 8.8).

Applicazioni della Genomica microbica

Lo studio del genoma microbico può avere grande importanza sia in campo clinico che biotecnologico. L'enorme biodiversità microbica, ancora largamente

inesplorata, costituisce un'inesauribile riserva di attività enzimatiche per le più svariate applicazioni biotecnologiche. In campo clinico, l'analisi genomica, oltre a contribuire all'identificazione delle caratteristiche genetiche responsabili della virulenza, può agevolare in modo determinante l'azione di contrasto contro le patologie microbiche attraverso lo sviluppo di antibiotici di nuova generazione e vaccini.

5.3 Genomi eucariotici

Struttura e organizzazione dei genomi eucariotici

Il sequenziamento completo di un gran numero di genomi eucariotici ha permesso di ottenere un quadro abbastanza dettagliato circa la loro struttura e organizzazione.

In realtà, i primi studi sulla quantità di DNA e quindi sulle dimensioni dei genomi degli organismi eucariotici e procariotici precedono la risoluzione della struttura del DNA avvenuta nel 1953 ad opera di Watson e Crick.

Nel 1948 Roger e Colette Vendrely mostrarono che tutte le cellule di una stessa specie animale possedevano una quantità costante di DNA, che venne per questo definita **valore C**. Fu subito evidente la non correlazione tra il valore C e la complessità dell'organismo, sia perché si osservava una grande gamma di dimensioni per organismi appartenenti a uno stesso raggruppamento tassonomico (ad es. protisti) che per il fatto che organismi apparentemente meno complessi (ad es. alghe) erano caratterizzati da un valore C molto superiore a quello delle cellule umane. Questa singolarità, che fu definita **paradosso del valore C**, venne risolta con la scoperta della frazione non codificante del genoma.

Nella **tabella 5.2** sono mostrati alcuni dati numerici sulla dimensione del genoma per alcune specie di organismi di diversi gruppi filogenetici, da *E. coli* a *Homo sapiens*, espressi sia come grandezza fisica del genoma (numero di pb e mm lineari del genoma idealmente steso nella sua lunghezza) che come numero di geni codificanti. In **figura 5.5** è invece riportata la gamma delle dimensioni dei genomi per alcune classi di organismi eucariotici, confrontata con quella dei genomi procariotici

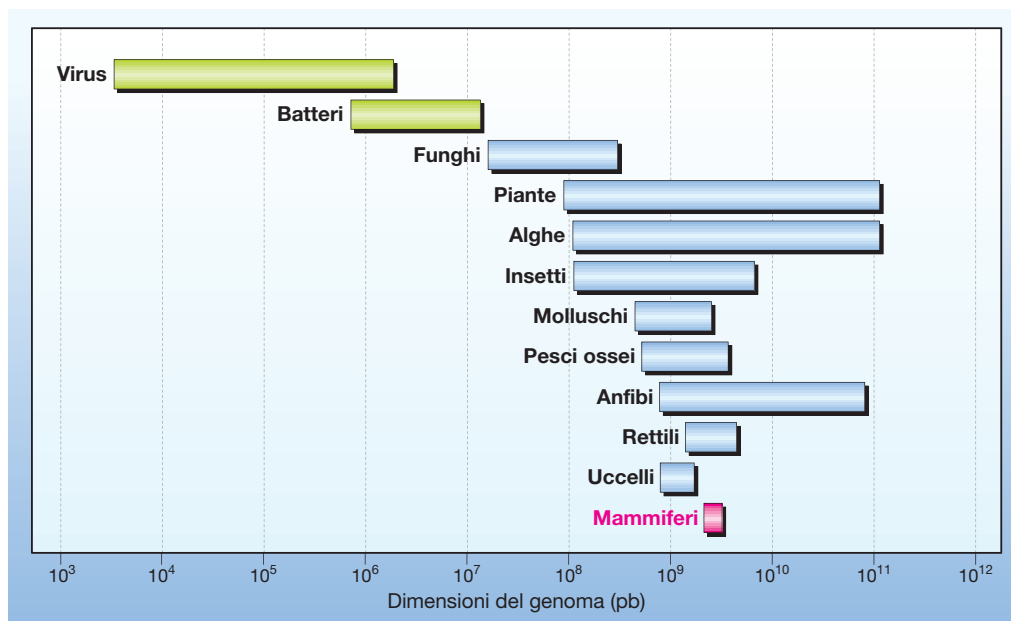


Figura 5.5 Gamma delle dimensioni dei genomi virali, procariotici ed eucariotici.

Tabella 5.2 Dimensioni e relativo numero stimato di geni dei genomi di alcuni organismi modello.

Genoma	Dimensioni		Numero di geni codificanti proteine
	Mpb	Lunghezza (cm)	
ΦX174 (piccolo fago)	0,005	0,00016	11
T7 (grande fago)	0,08	0,00259	150
Mimivirus (virus gigante)	1,2	0,03886	1000
<i>Escherichia coli</i> (procariote)	4,6	0,14895	4300
<i>Saccharomyces cerevisiae</i> (lievito)	13,5	0,43714	6000
<i>Caenorhabditis elegans</i> (nematode)	100	3,23810	21700
<i>Arabidopsis thaliana</i>	120	3,88571	26000
<i>Arabidopsis thaliana</i> (cloroplasto)	0,155	0,00502	87
<i>Drosophila melanogaster</i>	165	5,34286	12000
<i>Homo sapiens</i>	3300	106,857	23000
<i>Homo sapiens</i> (mitocondrio)	0,016	0,00052	37

e virali. Si può osservare come per alcune classi di organismi lo spettro di variabilità sia molto ampio (ad es. alghe e piante) e dovuto prevalentemente alla frazione non codificante del DNA, costituita essenzialmente da sequenze ripetute.

Al contrario dei genomi procariotici, che sono prevalentemente costituiti da una singola molecola di DNA circolare, i genomi eucariotici sono costituiti da DNA in forma lineare distribuito su un numero variabile di cromosomi. Non vi è tuttavia alcuna correlazione tra le dimensioni del genoma e il numero di cromosomi, o tra questo e la complessità dell'organismo.

Per esempio, l'eucariota unicellulare *Saccharomyces cerevisiae* ha un genoma di circa 13 Mpb organizzato in 16 cromosomi, mentre *Drosophila melanogaster* ha un genoma di circa 180 Mpb suddiviso in quattro cromosomi.

Caratteristiche composizionali dei genomi eucariotici

Circa la composizione in basi, i genomi eucariotici mostrano una minore variabilità del contenuto in G+C rispetto ai genomi procariotici, anche se differenze si possono osservare sia all'interno che tra i diversi raggruppamenti tassonomici.

Anche i genomi eucariotici, come abbiamo già visto per quelli procariotici, mostrano una marcata simmetria composizionale. Rispetto ai genomi procariotici, negli eucarioti si osserva una più marcata variazione intragenomica della composizione in basi. In particolare, nei vertebrati omeoterme come i mammiferi e gli uccelli, il genoma è costituito da un mosaico di segmenti di DNA, denominati **isocore**, di dimensioni $>> 300$ kpb, caratterizzati da una specifica e omogenea composizione in basi. Nell'uomo e negli altri mammiferi si osservano cinque diverse classi di isocore denominate L1, L2, H1, H2 e H3, caratterizzate da un contenuto crescente di G+C. Le isocore L1 e L2 (*Light*) sono povere in G+C e costituiscono oltre il 60% dell'intero genoma. Le isocore H1, H2 e H3 (*Heavy*) sono più ricche in G+C e contengono la maggior parte dei geni (fig. 5.6). In particolare le isocore H2 e H3 hanno una densità genica molto elevata.

Le isocore sono state scoperte mediante esperimenti in gradiente di densità di DNA genomico blandamente frammentato e la loro esistenza è stata poi confermata dal sequenziamento genomico dell'uomo e di altri organismi (fig. 5.7). Le isocore pesanti corrispondono a regioni del genoma caratterizzate da un minore livello di metilazione, oltre che da una struttura più aperta della cromatina e dalla prevalenza di geni housekeeping (costitutivamente espressi).

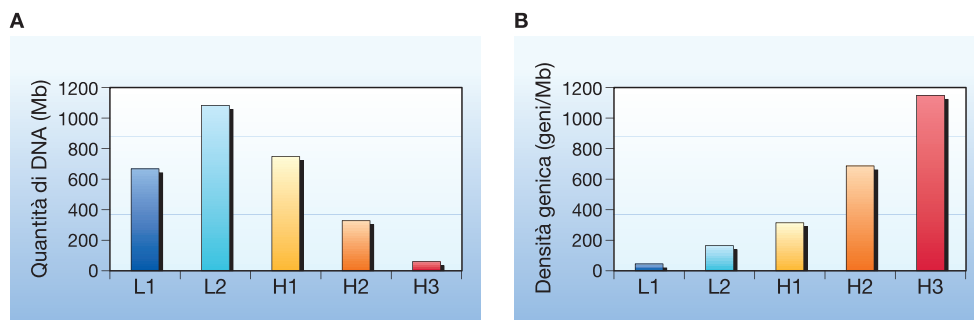


Figura 5.6 Abbondanza (A) e densità genica (B) di ciascuna delle cinque classi di isocore nel genoma umano.

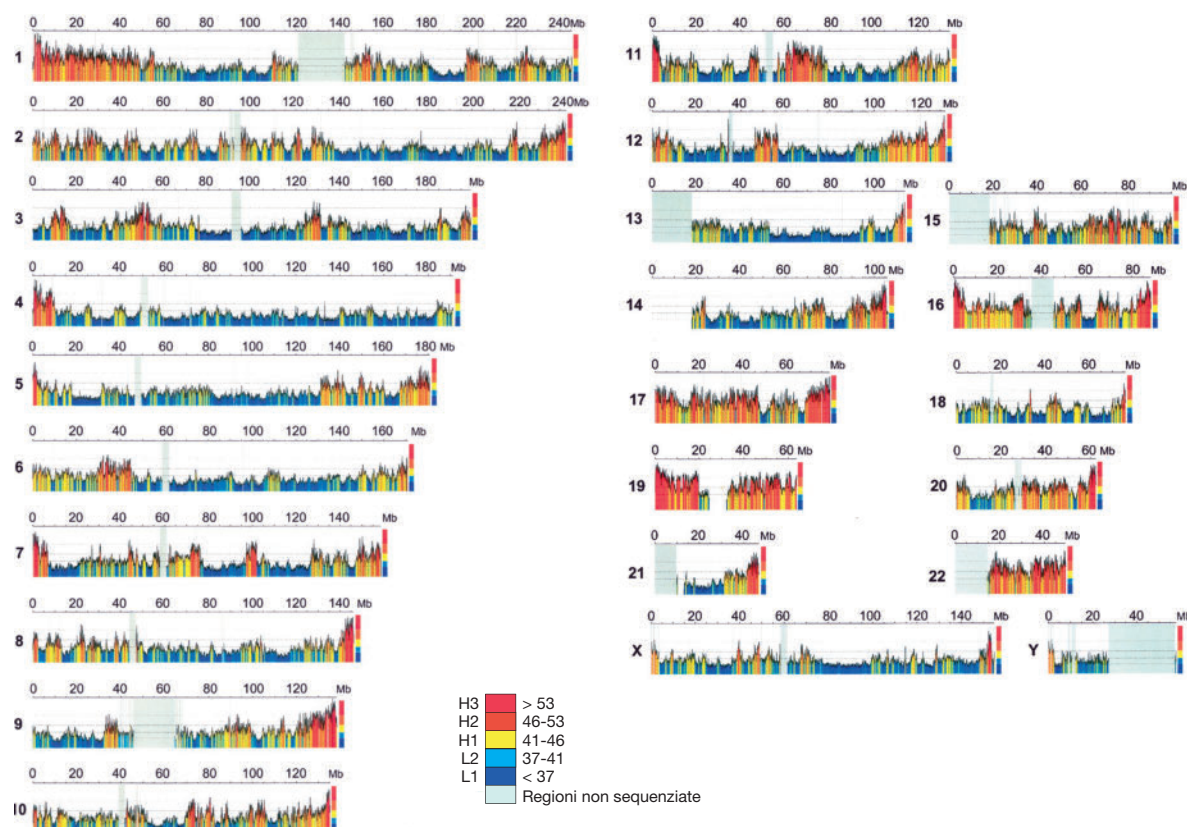


Figura 5.7 Struttura a isocore del genoma umano. È riportato il livello di GC% dei cromosomi umani (calcolati su finestre di 100 kpb). [Fonte: Costantini, M. et al. (2006), *Genome Res.*, 16, pp. 536-541.]

Contenuto genico dei genomi eucariotici

Una delle osservazioni più sorprendenti emerse dal sequenziamento del genoma dell'uomo e di molti altri organismi è che il numero dei geni è sostanzialmente conservato, indipendentemente dalle differenze, anche molto marcate, nelle dimensioni del genoma (**fig. 5.8**). Per esempio, il genoma umano è circa 250 volte più grande del genoma di lievito, ma il numero di geni codificanti proteine è dello stesso ordine di grandezza, poco più di 20 000 nell'uomo e circa 6000 nel lievito. All'interno dei metazoi (gli eucarioti multicellulari) le differenze nel nu-

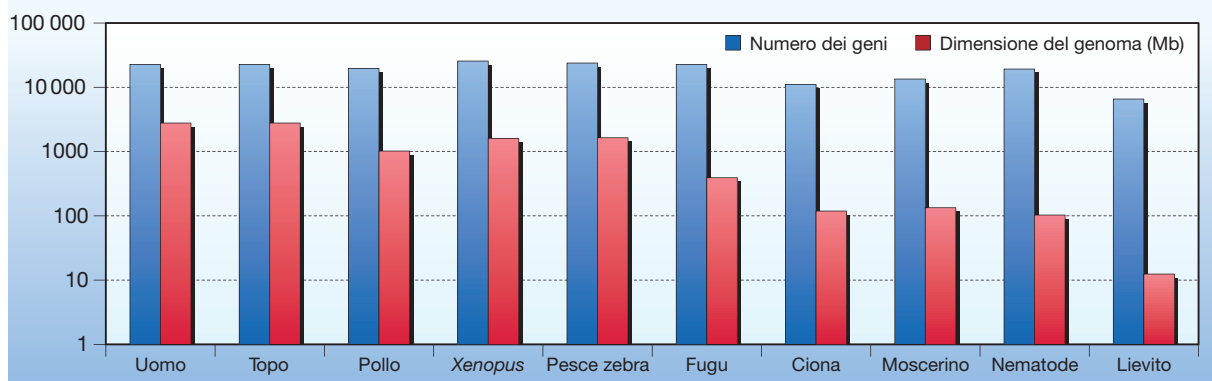


Figura 5.8 Assenza di correlazione tra numero di geni e dimensione del genoma negli eucarioti.

mero di geni sono ancora meno marcate. È quindi evidente che se il numero di geni è pressoché costante, le differenze nelle dimensioni del genoma sono sostanzialmente correlate alla frazione di DNA non codificante.

Caratteristiche dei geni eucariotici

I geni eucariotici presentano una grande varietà di strutture e dimensioni. La gamma di dimensioni si estende dalle poche decine di nucleotidi dei geni per tRNA fino a milioni di pb, come il gene umano per la distrofina, pari a 2,4 Mpb e la cui trascrizione completa richiede circa 16 ore.

Come vedremo in dettaglio nel capitolo 13 (*vedi* par. 13.1), i geni eucariotici, a differenza di quelli procariotici, sono generalmente discontinui, ovvero vengono espressi dall'apparato trascrizionale sotto forma di lunghi RNA precursori. Questi subiscono, nel nucleo, un processo di maturazione che comporta la rimozione di tratti polinucleotidici interni, denominati **introni**, e la concatenazione delle restanti sequenze, denominate esoni. L'mRNA maturo così prodotto sarà poi traslocato nel citoplasma, dove potrà essere tradotto dai ribosomi.

Mediamente, un gene umano contiene 8-9 introni, anche se alcuni geni sono del tutto privi di introni, come quelli per istoni, interferoni e ribonucleasi. All'opposto ci sono geni, come quello per la titina, con oltre 300 introni.

Le dimensioni degli esoni e degli introni sono estremamente variabili. La maggior parte degli esoni (non terminali) ha dimensioni comprese tra 50 e 250 nt. Tuttavia, si possono osservare microesoni di soli 6 nt, come pure esoni molto grandi, come il terzo esone del gene umano MUC16 che totalizza più di 20000 nt. Una variabilità ancora maggiore si osserva per gli introni, la cui lunghezza media nel genoma umano è di circa 3000 pb, ma con un intervallo di variabilità da poche decine di pb fino a oltre 1 Mpb. Sebbene vi sia un'elevatissima variabilità nella struttura dei geni umani, mediamente oltre il 90% della lunghezza degli RNA precursori è costituito da sequenze introniche (**fig. 5.9**).

Le proteine eucariotiche sono generalmente più grandi di quelle procariotiche. Per esempio, le proteine umane e quelle di *E. coli* hanno una lunghezza mediana di circa 375 e 275 amminoacidi, rispettivamente.

Le dimensioni più o meno compatte dei geni sono correlate al loro livello di espressione. È stato infatti osservato che i geni altamente espressi sono mediamente molto più compatti, con introni assenti o di piccole dimensioni.

Anche per i problemi prima evidenziati, a tutt'oggi non disponiamo del catalogo completo dei geni dell'uomo e degli altri organismi eucariotici di cui è stato completato il genoma.

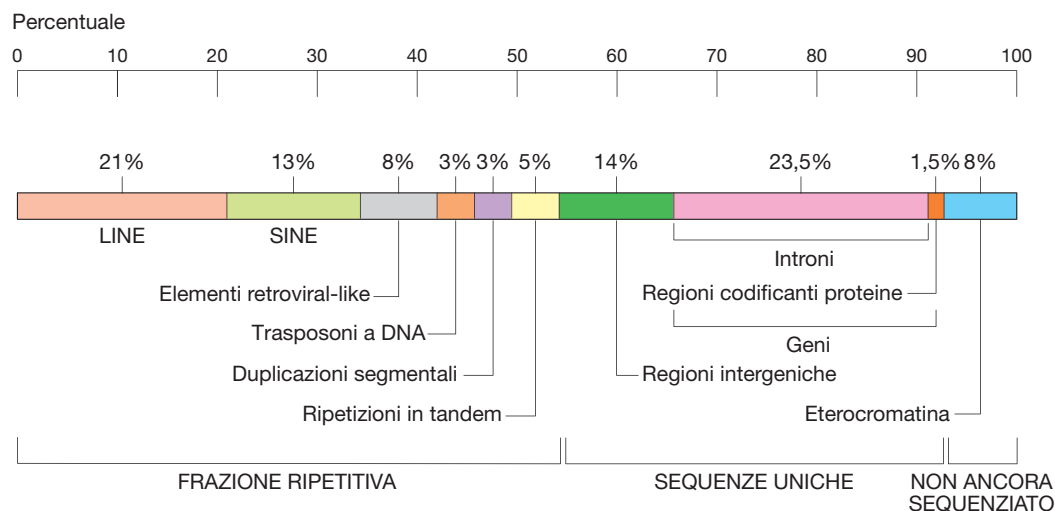


Figura 5.9 Composizione del genoma umano. Solo poco più dell'1% del genoma umano contiene l'informazione per la sintesi delle proteine. Tutto il resto è costituito da DNA non codifican-

te e, in particolare, da sequenze ripetute in tandem o intersperse che occupano oltre il 50% dell'intero genoma.

Secondo le stime più recenti, il genoma umano, che è quello su cui abbiamo più informazioni, contiene circa 23 000 geni codificanti proteine. La lunghezza totale degli esoni codificanti costituisce poco più dell'1% della porzione eucromatica dell'intero genoma (**fig. 5.9**). Sono annotati anche oltre 6000 geni per RNA non codificanti proteine, che includono rRNA, tRNA, snRNA, snoRNA, miRNA e altri ncRNA.

Ruolo degli introni nell'evoluzione dei geni eucariotici

La presenza degli introni è una caratteristica comune alla maggior parte dei geni degli organismi eucariotici. Anche se la presenza degli introni appare a prima vista un'inutile complicazione nel processo di espressione genica, questi costituiscono un'eccezionale fonte di variabilità che consente di esplorare ed eventualmente acquisire un numero pressoché infinito di soluzioni.

Vedremo più avanti come la struttura discontinua dei geni, costituiti da esoni e introni, favorisca l'origine di nuovi geni attraverso il processo dell'exon shuffling che genera nuovi assortimenti di esoni generalmente provenienti da geni diversi. Tale processo introduce variazioni strutturali nel genoma che sono poi fissate nella popolazione di individui di una specie o gruppo tassonomico attraverso un lungo processo evolutivo che può richiedere decine o centinaia di milioni di anni.

La struttura discontinua dei geni, che rende necessario il processo di splicing che nel nucleo genera i trascritti maturi, contribuisce in modo significativo all'espansione della capacità di espressione genica. Infatti, la maturazione del trascritto precursore attraverso lo splicing (*vedi* par. 13.1), responsabile del processo di "taglia e cuci" che rimuove gli introni e concatena gli esoni, può avvenire in molti modi differenti dando origine a trascritti e proteine alternativi a partire da uno stesso gene. Difatti, sebbene il numero di geni nei genomi procariotici ed eucariotici sia sostanzialmente dello stesso ordine di grandezza (**fig. 5.8**) il potenziale di espressione dei geni eucariotici, dotati di introni, è molto maggiore di quello dei geni procariotici (privi di introni), a causa della possibilità di "splicing alternativo" dei trascritti precursori dei primi. È stato osservato che lo splicing alternativo, di cui parleremo più estesamente nel capitolo 13 (*vedi* par. 13.1), riguarda la quasi totalità (> 95%) dei geni umani discontinui e comporta un'espansione della complessità del trascrittoma e del prote-

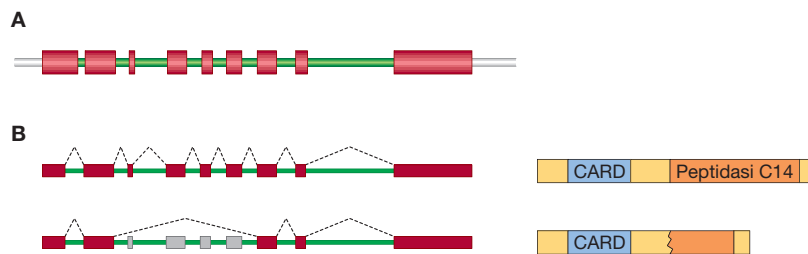


Figura 5.10 Splicing alternativo del trascritto del gene per la caspasi 9. Il gene per la caspasi 9 (A) dà origine, per splicing alternativo, a due diversi trascritti (B). Il primo codifica per una isoforma costitutiva denominata CASP9 (9 esoni, 416 aa) che induce apoptosi. Essa contiene un dominio CARD (*Caspase Recruitment*

Domain) e un dominio proteolitico (Peptidasi C14). Il secondo codifica per una isoforma più corta della proteina denominata CASP9S (5 esoni, 266 aa) e contiene un dominio CARD integro e un dominio tronco della Peptidasi C14. Questa seconda isoforma è priva dell'attività proteasica e agisce da inibitore dell'apoptosi.

oma di almeno un ordine di grandezza. Ciò implica che i circa 23 000 geni identificati nel genoma umano possano originare oltre 200 000 trascritti differenti. Questi possono presentare differenze sia nelle regioni non tradotte 5' e 3'UTR che nella regione codificante. Pertanto, le diverse forme di mRNA espresse da uno stesso gene possono produrre proteine differenti, che potrebbero anche avere funzioni distinte, perfino antagoniste (fig. 5.10). Analogamente, trascritti che differiscono nelle regioni 5' e 3'UTR possono essere sottoposti a una diversa regolazione post-trascrizionale, con conseguente diversa stabilità o efficienza traduzionale.

Lo splicing alternativo è dunque un meccanismo funzionalmente importante per l'individuo perché permette la produzione di più proteine implicate nello sviluppo e nel funzionamento dell'organismo. Come illustrato nel capitolo 13 (vedi par. 13.1), il processo di splicing è estesamente regolato e le sue disfunzioni sono responsabili di diverse malattie genetiche (spliceopatie, vedi anche finestra 13.1).

Alla fine degli anni settanta Walter Gilbert (fig. 5.11) ha avanzato per primo l'idea che la struttura discontinua esoni/introni dei geni, scoperta proprio in quegli anni da vari gruppi di ricerca, sia fondamentale anche per un differente processo, che chiamò **exon shuffling** (riassortimento degli esoni). Questo processo permette, nel corso dell'evoluzione, la nascita di nuove proteine. Queste infatti possono essere generate attraverso nuove combinazioni di esoni che possono realizzarsi se due geni si scambiano i rispettivi esoni formando nuovi geni caratterizzati da un nuovo assortimento di esoni rispetto a quello di partenza (fig. 5.12). Tale riassortimento è favorito in quei casi in cui un intero dominio proteico sia codificato da un singolo esone.

Il riassortimento degli esoni può aver luogo in seguito a crossing over diseguale, o illegittimo, che si può verificare nel caso di appaiamento non preciso tra due cromosomi omologhi durante la ricombinazione meiotica (fig. 5.12 e cap. 8). L'appaiamento impreciso può essere dovuto a regioni duplicate o a elementi ripetuti.

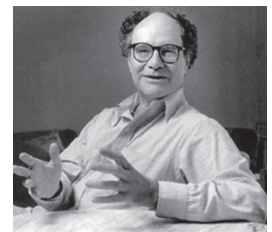


Figura 5.11 Walter Gilbert (1932-). Gilbert è un biochimico e fisico statunitense, premio Nobel per la chimica insieme con Frederick Sanger per il loro contributo alla determinazione delle sequenze di basi nel DNA. È stato il primo ad avanzare l'ipotesi che la struttura discontinua dei geni eucariotici favorisse la nascita di nuovi geni attraverso l'assortimento di nuove combinazioni di esoni, nel processo noto come *exon shuffling*.

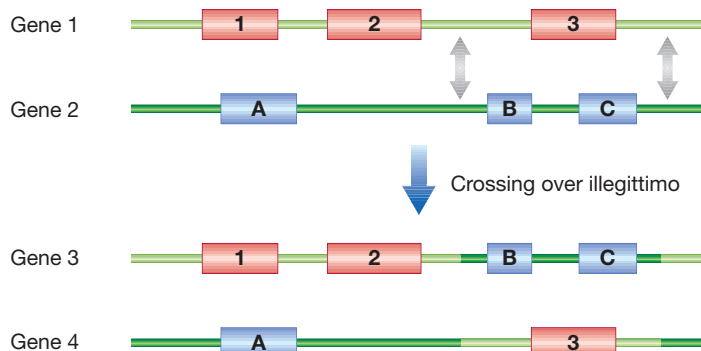
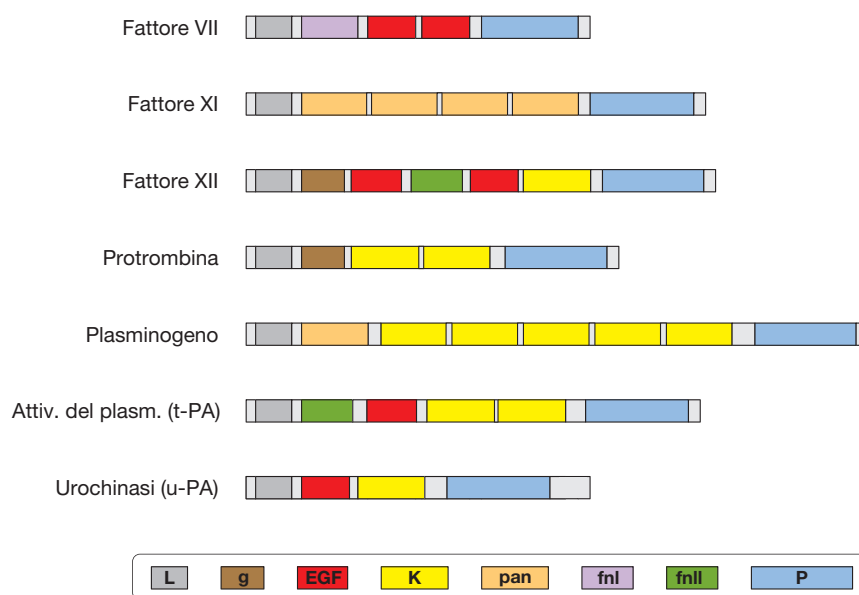


Figura 5.12 Riassortimento d'eson (exon shuffling) e formazione di nuovi geni. Due nuovi geni si possono formare attraverso il riassortimento dei rispettivi esoni (rappresentati dai box colorati) mediante crossing over illegittimo.

Figura 5.13 Origine di nuove proteine mediante *exon shuffling*. Nel corso dell'evoluzione si osserva che segmenti di DNA codificanti per moduli o domini nelle proteine vengono duplicati o riarrangiati. Il riassortimento di esoni tra diversi geni è alla base dell'origine delle famiglie geniche. Una di queste famiglie è rappresentata dai fattori di coagulazione del sangue i cui membri, rappresentati in figura, presentano domini simili in diverse combinazioni (L, peptide leader; g, dominio Gla; EGF, dominio EGF-like; K, dominio kringle; pan, dominio Pan; fnl, dominio della fibronectina di tipo I; fnII, dominio della fibronectina di tipo II; P, dominio proteasico).



La principale evidenza dell'intervento dell'exon shuffling nel processo evolutivo sono le proteine chimeriche o a mosaico. Queste sono proteine codificate da geni che contengono alcuni tratti che sono in comune con altri geni. Per esempio, i meccanismi di duplicazione, delezione o inserimento di esoni tipici dell'exon shuffling, possono spiegare l'origine evolutiva dei geni codificanti le proteine coinvolte nei processi di coagulazione del sangue e nella fibrinolisi in cui si osserva un diverso assortimento e composizione di specifici domini (fig. 5.13). Nuove varianti geniche possono anche essere originate in seguito all'inserimento di elementi mobili all'interno di introni, e loro successiva attivazione per l'acquisizione di nuovi siti di splicing. Questo processo, noto come **esonizzazione**, oltre che coinvolgere elementi trasponibili, particolarmente quelli di classe II (retrotrasposoni come Alu, LINE ecc.), può anche aver luogo in seguito alla retrotrasposizione all'interno di regioni introniche di cDNA originati da mRNA o loro porzioni (fig. 5.14).

La maggior parte dei processi che generano variazioni strutturali nel genoma, come quelli sopra descritti, sono potenzialmente deleteri e non vengono generalmente fissati nella popolazione in quanto subiscono un processo di selezione negativa. Per esempio, una traslocazione cromosomica comune nei pazienti con leucemia mieloide cronica genera un gene di fusione, noto come *bcr/abl*, i cui trascritti costituiscono un marcatore attendibile dello stato patologico.

Pseudogeni

Oltre ai geni funzionali, in grado cioè di esprimere uno o più trascritti che verranno poi tradotti in proteine, i genomi eucariotici contengono un gran numero di "falsi" geni, o **pseudogeni**. Gli pseudogeni sono geni inattivi non funzionali, incapaci cioè di esprimere proteine, che possono essere riconosciuti in base alla similarità di sequenza con i relativi geni funzionali.

Gli pseudogeni possono essere riconosciuti come tali sia per la presenza di mutazioni che determinano l'inserimento di codoni di stop prematuri, frameshift o estese delezioni/inserzioni, sia per l'assenza di un promotore funzionale che ne impedisce la trascrizione. In realtà, non appena un gene perde la sua attività funzionale comincia ad accumulare mutazioni "deleterie" che non sarebbe possibile osservare in un gene funzionale soggetto al controllo della selezione negativa.

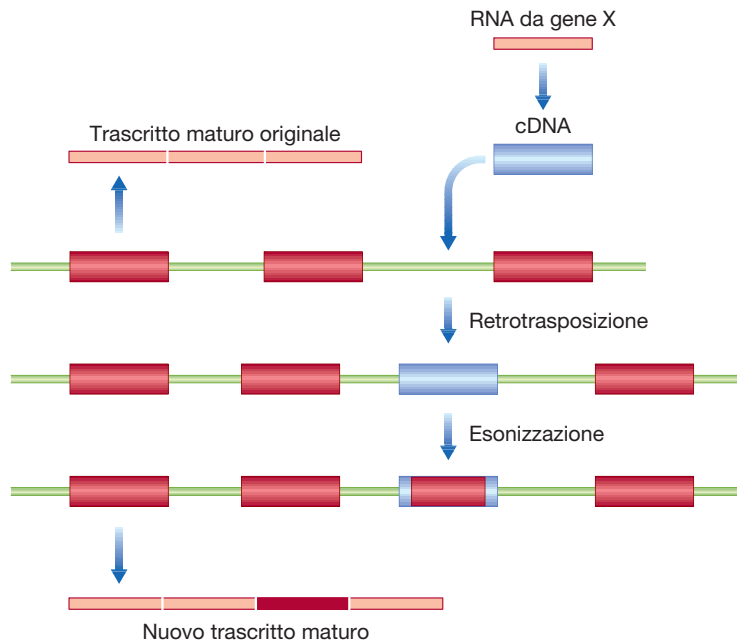


Figura 5.14 Schema che illustra il meccanismo di "esonizzazione". Il cDNA (completo o troncato) generato dalla trascrittasi inversa usando come stampo il trascritto di un qualunque gene o retrotrasposone (vedi oltre) si può integrare, attraverso un meccanismo di retrotrasposizione, nella regione intronica di un altro gene. Se la regione di DNA integrata contiene o acquisisce siti di splicing funzionali può dare origine a un nuovo esone non presente nel trascritto maturo originale.

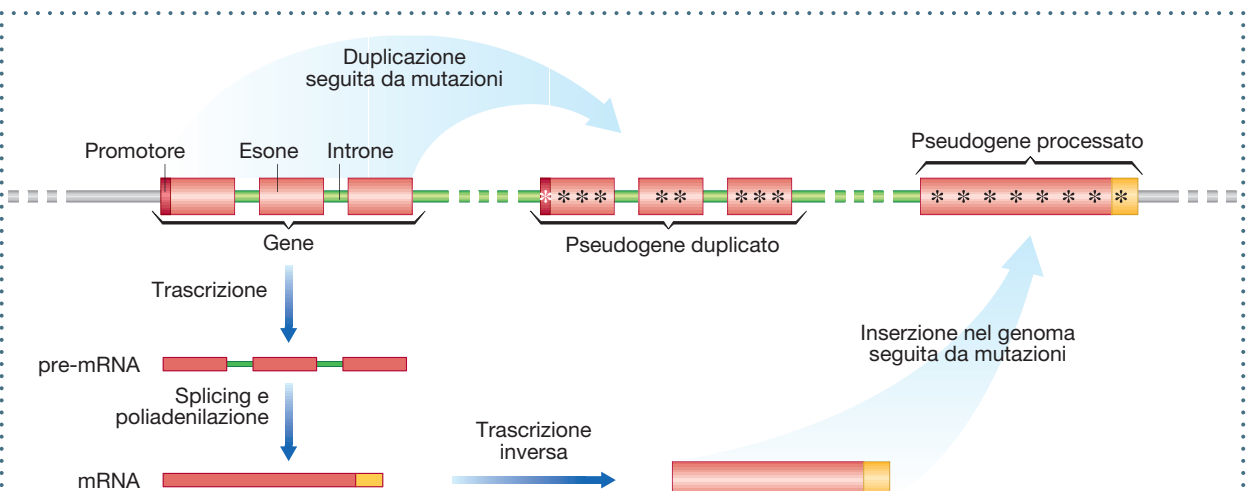


Figura 5.15 Origine degli pseudogeni. Uno pseudogene duplicato ha origine quando, in seguito a duplicazione genica, una seconda copia di un gene viene integrata in una nuova localizzazione genomica. Un pseudogene processato si forma, invece, a partire dall'RNA generato in seguito alla trascrizione del gene funzionale. L'mRNA maturo espresso da questo gene, contenente una coda di

poli(A) evidenziata in diverso colore, può essere retrotrascritto in cDNA e poi integrato nel genoma. Dopo l'integrazione gli pseudogeni di entrambe le classi cominciano ad accumulare mutazioni casuali (asterischi) che sfuggono al controllo della selezione negativa e ne consentono il riconoscimento.

Ma come si originano gli pseudogeni? Gli pseudogeni possono avere origine attraverso i due principali meccanismi molecolari responsabili della plasticità e del rimodellamento dei genomi nel corso dell'evoluzione: duplicazione e retrotrasposizione (fig. 5.15). La duplicazione genica può originare una nuova copia del gene che, originariamente, conserva la struttura in esoni e introni del gene originario e può rimanere inattiva se, per esempio, non viene integrata a valle di un promotore funzionale. Gli pseudogeni che invece si originano per retrotrasposizione sono riconoscibili per il fatto che sono privi di introni, mantengono delle vestigia della coda di poli-A e presentano alle estremità piccole ripetizioni dirette che marcano tipicamente le estremità degli inserti dopo la retrotrasposizione.

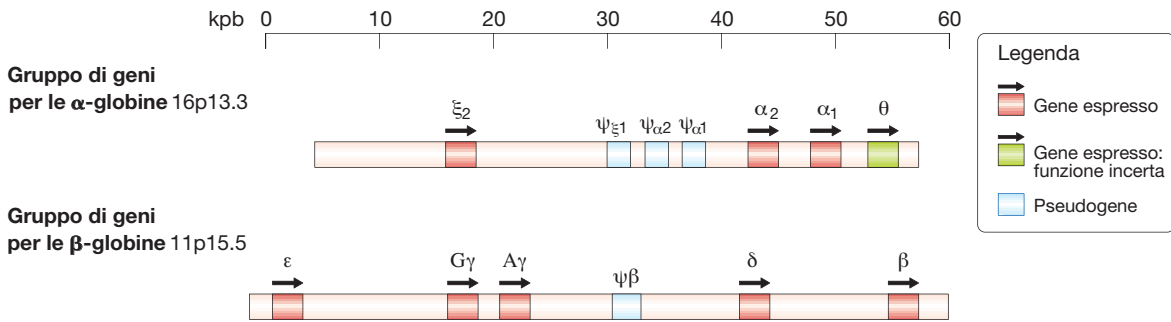


Figura 5.16 Struttura delle regioni del genoma umano contenenti i gruppi di geni dell' α - e β -globina. Si noti la presenza di pseudogeni duplicati in prossimità dei geni di origine (le frecce indicano la direzione della trascrizione).

A seconda del meccanismo utilizzato nella retrotrasposizione gli pseudogeni vengono classificati in **pseudogeni duplicati** e **pseudogeni processati**. Gli pseudogeni duplicati si trovano inizialmente vicini ai geni di origine (fig. 5.16), anche se poi possono traslocare in altre parti del genoma. Al contrario gli pseudogeni processati vanno a integrarsi a caso in qualsiasi parte del genoma. Gli pseudogeni processati derivano in genere da geni housekeeping o comunque espressi nelle cellule germinali.

Nel genoma umano sono stati identificati oltre 19 000 pseudogeni, dei quali oltre il 40% appartengono alla classe degli pseudogeni processati.

È stato osservato che geni altamente espressi, come quelli che codificano per le proteine ribosomali, possono dare origine a un maggior numero di copie di pseudogeni processati.

L'inattivazione di un gene, attraverso la degenerazione in pseudogene, può anche essere indipendente dall'evento di duplicazione o retrotrasposizione. Può infatti essere dovuta a mutamenti nella pressione selettiva su determinati caratteri nel corso dell'evoluzione. Per esempio, è noto che nell'uomo si è avuta una drastica riduzione nel numero di geni per i recettori dell'olfatto (OR, *Olfactory Receptors*), una delle famiglie di geni più numerosa, che nei mammiferi non primati consta di oltre un migliaio di membri. Circa la metà di questi si è trasformata in pseudogeni nell'uomo. Un'analoga situazione si osserva anche in altri primati che con l'uomo hanno in comune la capacità di visione tricolore. È evidente che il vantaggio conferito dall'acquisizione di questo carattere ha ridotto la pressione selettiva sul mantenimento delle capacità olfattive, che sono state ridimensionate nell'uomo e in altri primati.

Anche se gli pseudogeni sono per definizione geni inattivi, non è detto che siano del tutto privi di funzione. Infatti, una piccola parte di essi risulta essere trascritta, e l'RNA prodotto, anche se non codificante per proteine, potrebbe svolgere qualche ruolo funzionale. Per esempio, è stato osservato che il trascritto dello pseudogene NOS (*Nitric Oxide Synthase*) regola negativamente la traduzione dell'omologo trascritto funzionale. Inoltre, è stato osservato che l'inattivazione dello pseudogene Markorin 1 nel topo induce gravi disfunzioni dello sviluppo.

Alcuni pseudogeni possono anche "risorgere", ovvero essere riattivati nel corso del processo evolutivo. In questo modo potrebbero costituire una riserva di nuovi geni potenziali da riattivare in risposta a specifiche condizioni.

Sequenze ripetute dei genomi eucariotici

La presenza di sequenze ripetute nei genomi eucariotici è stata dimostrata, ancor prima che venissero introdotte le tecniche di sequenziamento, mediante lo studio della cinetica di riassociazione del DNA. Come vedremo nel capitolo 20 (vedi par. 20.1) sulle tecniche della Biologia Molecolare, i genomi procariotici presentano delle curve di riassociazione sigmoidali semplici che indicano l'assenza

Tabella 5.3 Numero di copie e organizzazione dei geni per rRNA in procarioti ed eucarioti.

Organismo	Numero di copie del gene per il precursore degli rRNA	Organizzazione genomica
<i>E. coli</i>	7	Disperse nel genoma
<i>S. cerevisiae</i>	140	Raggruppate in un locus nel cromosoma XII
<i>A. thaliana</i>	570	Raggruppate in 2 loci nei cromosomi 2 e 4
<i>D. melanogaster</i>	400	Raggruppate in 2 loci nei cromosomi X e Y
<i>X. laevis</i>	450	Raggruppate in un locus
<i>H. sapiens</i>	280	Raggruppate in 5 loci, nei cr. 13, 14, 15, 21 e 22

di apprezzabili frazioni di DNA ripetuto. Infatti, tranne pochissime eccezioni, i genomi procariotici non presentano sequenze ripetute. Anzi, se per duplicazione di un tratto di DNA si formano nel genoma due copie della stessa sequenza non molto distanti tra loro, queste tendono a ricombinare risultando nell'escissione di un tratto genomico e lasciando una sola delle due copie.

L'eccezione più notevole a questa regola è la presenza di più copie del gene che codifica per gli RNA ribosomali (rRNA). Nel genoma di *E. coli*, per esempio, ci sono 7 copie di questo gene, probabilmente per ragioni quantitative, considerando che gli rRNA, insieme ai tRNA, sono i trascritti di gran lunga più rappresentati nella cellula (**tab. 5.3**). Da notare che queste 7 copie del gene per l'rRNA non sono organizzate in tandem, come avviene invece per le multiple copie dei geni per gli rRNA negli eucarioti, ma sono ampiamente disperse lungo tutto il genoma, probabilmente per minimizzare la possibilità, di cui abbiamo sopra parlato, di ricombinazione e quindi di delezioni genomiche.

Al contrario, le complesse curve delle cinetiche di riassociazione a più gradini che si osservano per DNA di origine eucariotica (*vedi* fig. 20.5) dimostrano la presenza in questi genomi di distinte frazioni di DNA: una frazione di sequenze "uniche" (o quasi uniche), cioè non ripetute, una frazione di DNA mediamente ripetuto, con una ripetitività compresa tra una decina fino a migliaia di copie, e una frazione di DNA altamente ripetuto, fino a centinaia di migliaia o anche milioni di copie. La frazione di genoma a sequenze uniche, o non ripetitivo, contiene tra l'altro la quasi totalità dei geni veri e propri che codificano per proteine, che infatti sono presenti nel genoma in unica o pochissime copie. La frazione mediamente ripetuta comprende sequenze ripetute intersperse, trasposoni e retrotrasposoni attivi o non più attivi, ma anche alcuni geni perlopiù per RNA non codificanti, tra cui il precursore degli rRNA presente in centinaia di copie organizzate in gruppi in tandem. Infine la frazione di DNA altamente ripetuto è costituita da sequenze semplici o DNA satelliti.

La successiva messa a punto delle tecniche di sequenziamento del DNA e la determinazione delle sequenze nucleotidiche di interi genomi procariotici ed eucariotici hanno sostanzialmente confermato queste conclusioni. Per esempio, se consideriamo il genoma umano, osserviamo che oltre il 98% è costituito da DNA non codificante. La percentuale è molto simile negli altri mammiferi e nella maggior parte dei vertebrati. In particolare, la porzione non codificante del DNA è prevalentemente costituita da sequenze ripetute in tandem o intersperse che complessivamente occupano oltre il 50% dell'intero genoma umano (**fig. 5.9**).

DNA satellite o altamente ripetitivo

Quando una preparazione di DNA nucleare da organismi eucariotici viene ultracentrifugata in gradiente di CsCl (*vedi* par. 20.3), oltre alla banda principale che

Figura 5.17 DNA satellite. DNA satelliti del genoma di *Drosophila melanogaster* (A) e *Mus musculus* (B) individuati mediante centrifugazione isopicnica su cloruro di cesio. La densità di galleggiamento delle varie frazioni, la cui abbondanza relativa è proporzionale all'area sotto i picchi relativi, è funzione del contenuto in G+C.

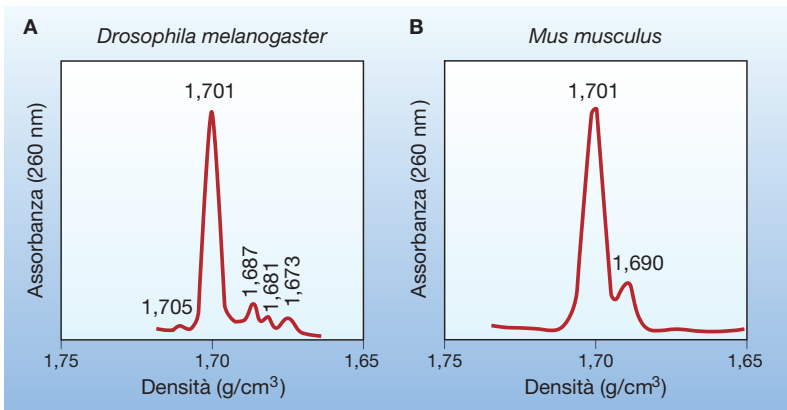


Tabella 5.4 Sequenze semplici ripetute nel DNA satellite di *Drosophila melanogaster*.

Sequenza	% del genoma	Copie/genoma
5'-AAGAG-3' 3'-TTCTC-5'	5,6	$2,5 \times 10^6$
5'-AATAT-3' 3'-TTATA-5'	3,1	$1,4 \times 10^6$
5'-AAGAC-3' 3'-TTCTG-5'	2,4	$1,1 \times 10^6$

rappresenta il grosso del DNA genomico, si osservano anche una o più bande minori, accessorie, più dense o meno dense del DNA principale, che per questo vengono chiamate **DNA satelliti** (fig. 5.17). I DNA satelliti, che costituiscono sostanzialmente la frazione altamente ripetitiva del genoma, vengono chiamati anche **DNA a sequenza semplice**. Infatti, il sequenziamento di alcuni DNA satelliti ha messo in evidenza che essi sono in genere costituiti da numerose (da decine di migliaia a milioni) ripetizioni in tandem di una breve sequenza, in blocchi (gruppi) distribuiti in vari siti e in vari cromosomi. I DNA satelliti in genere non sono trascritti e si trovano nelle regioni eterocromatiche dei cromosomi. La ragione per cui queste sequenze formano dei satelliti in gradiente di CsCl è che a causa della loro brevità presentano spesso una composizione in basi che devia da quella del grosso del DNA genomico, il quale, negli organismi eucariotici, ha circa il 40% di G+C. Per esempio, il DNA satellite di topo è costituito da varianti della sequenza di nove nucleotidi GAAAAANGT, quindi ha una composizione in basi del 22% G+C rispetto al 40% del grosso del DNA, dal quale infatti si separa in gradiente di CsCl come una banda satellite leggera. In qualche caso, per pura coincidenza, la sequenza monomerica ha una composizione in basi uguale o molto simile a quella del DNA totale e quindi non si separa come banda individuale nel gradiente di CsCl. In questo caso viene chiamato **satellite criptico**.

Illustreremo ora alcuni casi specifici per dare un'idea di come sono strutturati ed evolvono i DNA satelliti. In *Drosophila melanogaster* sono stati osservati vari satelliti (fig. 5.17A) che corrispondono a diverse sequenze semplici molto brevi e altamente ripetute in tandem, di cui le più rilevanti sono mostrate in **tabella 5.4** (in realtà la situazione è un po' più complessa per la presenza di altre sequenze semplici, somiglianti a queste ma ripetute meno volte). Come si può notare, le tre sequenze mostrate in tabella presentano una certa somiglianza, suggerendo che probabilmente si sono originate da una sequenza ancestrale comune. Esse rappresentano nell'insieme oltre l'11% del genoma, pari a circa 5×10^6 ripetizioni. In questo caso le diverse copie di ciascun blocco di sequenze sono molto conservate con pochissime variazioni. Se consideriamo altre specie di *Drosophila* troviamo

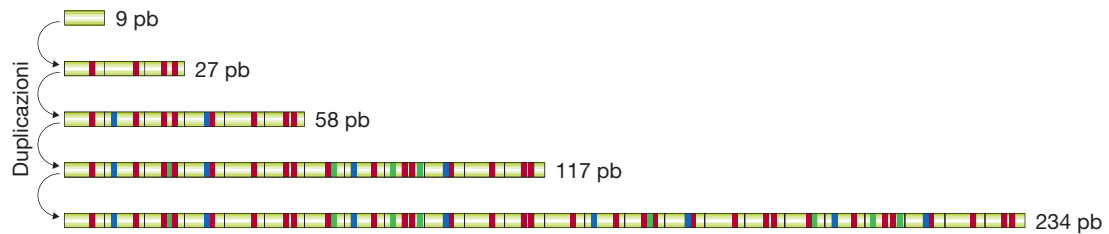


Figura 5.18 Evoluzione del DNA satellite centromerico di topo. L'unità ripetitiva di 234 pb si sarebbe originata dalla sequenza di 9 pb attraverso una serie di duplicazioni in tandem alternate a mutazioni (sostituzioni o inserzioni) in una o più delle copie, se-

condo il seguente schema: (1) triplicazione della sequenza da 9 pb; 4 mutazioni (in rosso nella figura); (2) duplicazione; 2 mutazioni per inserzione (blu); (3) duplicazione; 4 mutazioni (verde); (4) duplicazione.

situazioni analoghe ma con sequenze e livelli di ripetitività diversi. La situazione è un po' diversa nel topo e negli altri mammiferi, incluso l'uomo.

Nel genoma di topo, per esempio, è presente un DNA satellite (**fig. 5.17B**), molto ben studiato, che si trova localizzato nelle regioni centromeriche di tutti i cromosomi. Questo satellite presenta una sequenza precisamente ripetuta di 234 pb. Andando però a osservare con attenzione questa unità ripetitiva fondamentale ci si accorge che essa è costituita dalla ripetizione non precisa di sequenze più brevi. In particolare si può dividere la sequenza di 234 pb in due sequenze quasi uguali di 117 pb, ciascuna delle quali è costituita da due sequenze di 58 pb un po' più divergenti. Proseguendo in questa maniera si arriva a identificare una sequenza di 9 pb di lunghezza che sarebbe alla base dell'evoluzione di questo DNA satellite. Il consenso delle varie copie divergenti di questa sequenza è GAAAAA(T/A/C)GT. Analizzando la distribuzione delle varie copie divergenti della sequenza si è potuto risalire alla storia evolutiva di questo DNA satellite, illustrata in **figura 5.18**. La sequenza da 234 pb così originata avrebbe poi subito una grande espansione che l'ha portata agli attuali alti livelli di ripetitività nel genoma di topo.

Nella specie umana sono stati identificati vari satelliti, per lo più localizzati nelle regioni centromeriche, ma anche in altre regioni eterocromatiche. Tra questi ricordiamo la famiglia dei DNA **satelliti α** (chiamati anche **alphoid DNA**), la cui unità monomerica è lunga 171 pb e che, analogamente al DNA satellite di topo, sarebbe il risultato della duplicazione e divergenza di sequenze più corte. Il DNA alfoide forma dei blocchi che rappresentano una porzione cospicua del genoma umano (> 60 Mpb), ma che sono ancora poco rappresentati nelle sequenze genomiche prodotte nell'ambito dei progetti di sequenziamento. Satelliti della famiglia alfa sono presenti in tutti i primati del vecchio continente.

Da queste e altre osservazioni possiamo concludere che l'evoluzione dei DNA satelliti è il risultato di occasionali eventi di successive duplicazioni in tandem (chiamate alle volte repliche saltatorie), che risultano in un improvviso e rapido aumento del numero di copie ripetute, alternate con eventi mutazionali che introducono una certa divergenza tra le copie. Blocchi di sequenze ripetute possono essere anche facilmente persi per escissione conseguente a ricombinazione tra diverse copie della sequenza. Questo meccanismo di espansione, mutazione e perdita delle sequenze del DNA altamente ripetute fa sì che, evolutivamente, i DNA satelliti cambino molto rapidamente, sia come sequenza nucleotidica che come livello di ripetitività. Infatti, specie affini hanno in genere DNA satelliti ben distinguibili, tanto che queste sequenze vengono anche utilizzate come marcatori per risolvere casi di dubbia appartenenza sistematica.

DNA mediamente ripetitivo: microsatelliti e minisatelliti

Una frazione del genoma eucariotico è costituita dai cosiddetti **minisatelliti** e **microsatelliti**, che però sono solo in parte da accomunare con i DNA satelliti. A

FINESTRA 5.1

IL TEST DEL DNA

Ogni individuo eredita il proprio patrimonio genetico per il 50% dal padre e per il 50% dalla madre (con l'eccezione del DNA mitocondriale ereditato esclusivamente per via materna). Tranne che nei gemelli omozigoti che condividono lo stesso patrimonio genetico, ogni individuo risulta unico, e anche differente sia dal padre che dalla madre. Questa caratteristica ha reso il DNA uno strumento ampiamente utilizzato per gli studi genealogici (ad es. test di paternità) o per individuare i responsabili di azioni criminali. Per il test del DNA viene utilizzata l'analisi dei **polimorfismi del DNA microsatellite (STR, Short Tandem Repeats)**. Questi sono polimorfismi di lunghezza, in quanto varia il numero di copie ripetute da individuo a individuo.

Il DNA può essere estratto da qualunque tessuto umano. Per esempio i tamponi buccali, simili a grossi "cotton fioc" che passati sulle pareti interne delle guance raccolgono le cellule di sfaldamento dell'epitelio boccale, vengono generalmente usati per eseguire il test di paternità. Le principali fonti di DNA sulla scena di un crimine sono invece costituite da sangue, liquido seminale, capelli, pelle ecc.

Il DNA estratto viene poi amplificato mediante la tecnica PCR (vedi par. 20.6) utilizzando una miscela di primer, corrispondenti alle regioni uniche che fiancheggiano le sequenze ripetute, che consentono di analizzare simultaneamente più loci STR.

Ogni locus STR ha un elevato numero di alleli, distinti tra loro in base al numero di ripetizioni. Ogni cromosoma può avere due alleli identici, ovvero con lo stesso numero di ripetizioni (omozigosità), o diversi, con differente numero di ripetizioni (eterozigosità). La caratterizzazione genetica di molti loci di uno stesso individuo rappresenta il suo **profilo genetico (DNA fingerprint)**. Nel sistema CoDIS (*Combined DNA Index System*) utilizzato dall'FBI vengono analizzati simultaneamente 13 loci STR. I dati vengono poi immagazzinati in una banca dati in modo da riconoscere eventuali sospetti. Con questo sistema vi è una sola possibilità su un milione di miliardi che due individui posseggano lo stesso profilo genetico. La maggior parte dei kit in commercio consente anche la determinazione del sesso mediante lo studio dell'**amelogenina**, un gene codificante una proteina presente nello smalto dei denti che mostra variazioni tra maschi e femmine: la variante del gene nel cromosoma X è caratterizzata da una delezione di 6 pb, non presente nella variante legata al cromosoma Y. Tipicamente questi kit hanno molecole fluorescenti legate ai primer in modo da poter analizzare simultaneamente più loci con prodotti di amplificazione di dimensioni simili. I prodotti di amplificazione vengono separati per elettroforesi capillare e la forma allelica di ogni locus determinata attraverso il confronto con uno standard di riferimento.



Figura 5F1.1 DNA fingerprinting per il test del DNA. Analisi di tre loci STR di un individuo caratterizzati attraverso il confronto con

uno standard. Si noti come il locus D3S1358 risulta omozigote con un valore di 15, mentre i loci vWA e FGA sono entrambi eterozigoti.

differenza dei satelliti, queste sequenze sono ripetute nel genoma un numero molto limitato di volte e quindi non fanno parte della frazione altamente ripetitiva del DNA. I prefissi mini- e micro- stanno a indicare la brevità della sequenza ma anche il basso numero di ripetizioni.

I microsatelliti sono costituiti da monomeri di dimensioni variabili da 1 a 10 pb, ripetuti in tandem circa 10-30 volte. Essi sono anche detti **elementi SSR**

Tabella 5.5 Patologie associate all'espansione di triplette. In alcuni casi si osservano livelli intermedi di espansione (pre) che non hanno un fenotipo patologico ma che possono essere trasmessi alla progenie. (Fonte: Cummings e Zoghbi, ARGHG 2000.)

Malattia	Gene mutante	Locus	Proteina	Sequenza ripetuta			
				Ripetizione	Normale	Malattia	Sito di ripetizione
Sindrome dell'X fragile	FMR1 (FRAXA)	Xq27.3	FMR-1 (FMRP)	CGG	6-53	60-200 (pre) > 230	5'UTR
Sindrome dell'XE fragile	FMR2 (FRAXE)	Xq28	FMR-2	GCC	6-35	60-200 (pre) > 200	5'UTR
Atassia di Friedreich	X25	9q13-21.1	Fratassina	GAA	7-34	34-80 (pre) > 100	Introne 1
Distrofia miotonica	DMPK	19q13	DMPK (<i>Myotonic Dystrophy Protein Kinase</i>)	CGT	5-37	50-migliaia	3'UTR
Atrofia muscolare spinale e bulbare (malattia di Kennedy)	AR	Xq13-21	Recettore degli androgeni (AR)	CAG	9-36	38-62	Regione codificante
Malattia di Huntington	HD	4p16.3	Huntington	CAG	6-35	49-88	Regione codificante
Atrofia dentato-rubro-pallidoluisiana (sindrome di Haw River)	DRPLA	12p13.31	Atrofina-1	CAG	6-35	49-88	Regione codificante
Atassia spinocerebellare di tipo 1	SCA1	6p23	Atassina-1	CAG	6-44	39-82	Regione codificante
Atassia spinocerebellare di tipo 2	SCA2	12q24.1	Atassina-2	CAG	15-31	36-63	Regione codificante
Atassia spinocerebellare di tipo 3 (malattia di Machado-Joseph)	SCA3 (MJD1)	14q32.1	Atassina-3	CAG	12-40	55-84	Regione codificante
Atassia spinocerebellare di tipo 6	SCA6	19p13	α 1A subunità voltaggio-dipendente del canale del calcio	CAG	4-18	21-33	Regione codificante
Atassia spinocerebellare di tipo 7	SCA7	3p12-13	Atassina 7	CAG	4-35	37-306	Regione codificante
Atassia spinocerebellare di tipo 8	SCA8	13q21	Nessuna	CGT	16-37	da 110 a < di 250	3'UTR
Atassia spinocerebellare di tipo 12	SCA12	5q31-33	PP2A-55 β (PPP2R2B)	CAG	7-28	66-78	5'UTR

(*Simple Sequence Repeats*) o **STR** (*Short Tandem Repeats*). Loci di microsatelliti sono dispersi in moltissime localizzazioni genomiche nei diversi cromosomi.

I minisatelliti sono costituiti da unità monomeriche di dimensioni variabili, comprese tra 11 e 100 pb, che formano blocchi più estesi di lunghezza fino a 20 kpb.

I micro- e minisatelliti mostrano un elevato grado di polimorfismo nella popolazione umana in quanto il numero di ripetizioni all'interno di uno specifico locus può variare da individuo a individuo a causa di un meccanismo noto come **sci-**

volamento (*slippage*) della replicazione. Per questo vengono usati come marcatori dell'individualità genetica nel cosiddetto "test del DNA" che viene oggi normalmente utilizzato dalla polizia scientifica per scoprire gli autori dei crimini (**finestra 5.1**).

Anomale espansioni delle ripetizioni, particolarmente quelle costituite da triplette di nucleotidi, localizzate sia in regioni codificanti che non codificanti, sono associate a diverse patologie (**tab. 5.5**).

Sequenze di DNA ripetitivo intersperse nel genoma

Le sequenze ripetute intersperse corrispondono a elementi mobili o trasponibili in grado di "spostarsi" da una posizione all'altra del genoma. Esse sono costituite da: (1) elementi mobili di classe I o **retrotrasposoni**; (2) elementi mobili di classe II o **trasposoni a DNA** (**fig. 5.19**).

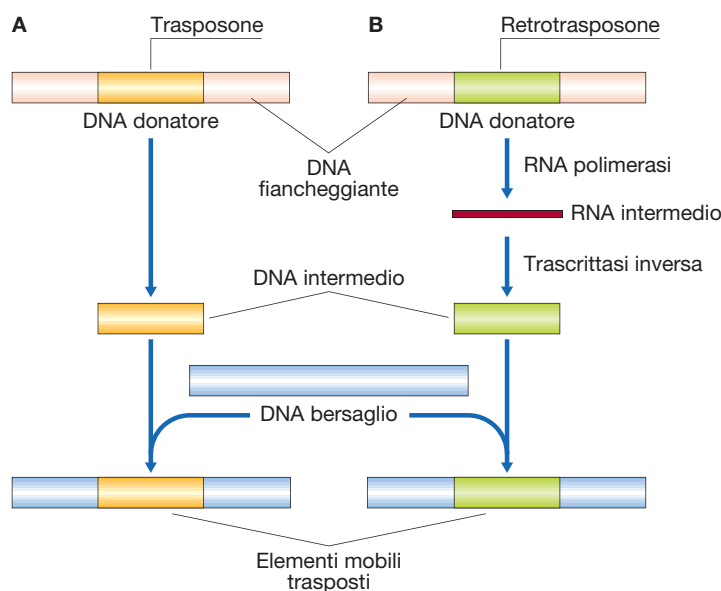
I retrotrasposoni, che vengono originati dall'attività della trascrittasi inversa a partire da un intermedio a RNA, possono essere suddivisi a loro volta in: **elementi LTR** (*Long Terminal Repeats*), caratterizzati dalla presenza di lunghe sequenze terminali ripetute di dimensioni variabili da poche centinaia fino a oltre 5 kpb, cui appartengono anche i retrovirus; **elementi non-LTR**, che comprendono gli elementi LINE (*Long Interspersed Nuclear Elements*) e SINE (*Short Interspersed Nuclear Elements*).

Infine, tutti gli elementi mobili sopra descritti possono essere distinti in "autonomi" e "non autonomi" a seconda che siano o non siano in grado di eseguire autonomamente il processo di trasposizione, ovvero che questo richieda attività enzimatiche codificate dagli elementi autonomi. Per esempio, la trascrittasi inversa codificata dagli elementi LINE è necessaria per la trasposizione di elementi SINE.

Gli elementi mobili interspersi sono ubiquitari in tutti i genomi eucariotici, di cui costituiscono una frazione cospicua (**tab. 5.6**). Essi sono importanti modulatori della plasticità del genoma e giocano un ruolo primario nella generazione di nuove specie, possono contribuire a introdurre elementi di regolazione dell'espressione genica quali enhancer e/o siti di terminazione della trascrizione, e attraverso un processo di esonizzazione possono introdurre nuovi segmenti peptidici nelle regioni codificanti dei geni. Si stima che circa il 4% della frazione codificante del genoma umano derivi da elementi trasponibili.

I trasposoni a DNA, che quindi non utilizzano un intermedio a RNA, sono caratterizzati da un meccanismo di propagazione conservativo o replicativo, a

Figura 5.19 Meccanismo di trasposizione degli elementi di classe I (B) e classe II (A).



seconda che la trasposizione lasci immutato o raddoppi il numero di copie dell'elemento trasposto nel genoma (**fig. 5.20**). I trasposoni a DNA contengono generalmente una sequenza codificante per la trasposasi, un enzima che si lega all'estremità del trasposone, mobilizzandolo in un'altra regione genomica o attraverso un meccanismo di trasposizione conservativa, di tipo "taglia e cuci" in cui rimane costante il numero di copie, oppure attraverso un meccanismo replicativo in cui la copia originale rimane al suo posto. I membri più noti di questa classe sono gli elementi *Ac* e *Ds* del granturco, identificati negli anni quaranta del secolo scorso da Barbara McClintock (*vedi* par. 8.3) e gli elementi P di *Drosophila*.

In **figura 5.21** è illustrata schematicamente la struttura dei principali elementi retrotrasponibili. Gli elementi LTR o retrotrasposoni virali, di dimensioni pari a 6-9 kpb, presentano analogie con i retrovirus e lunghe ripetizioni dirette alle estre-

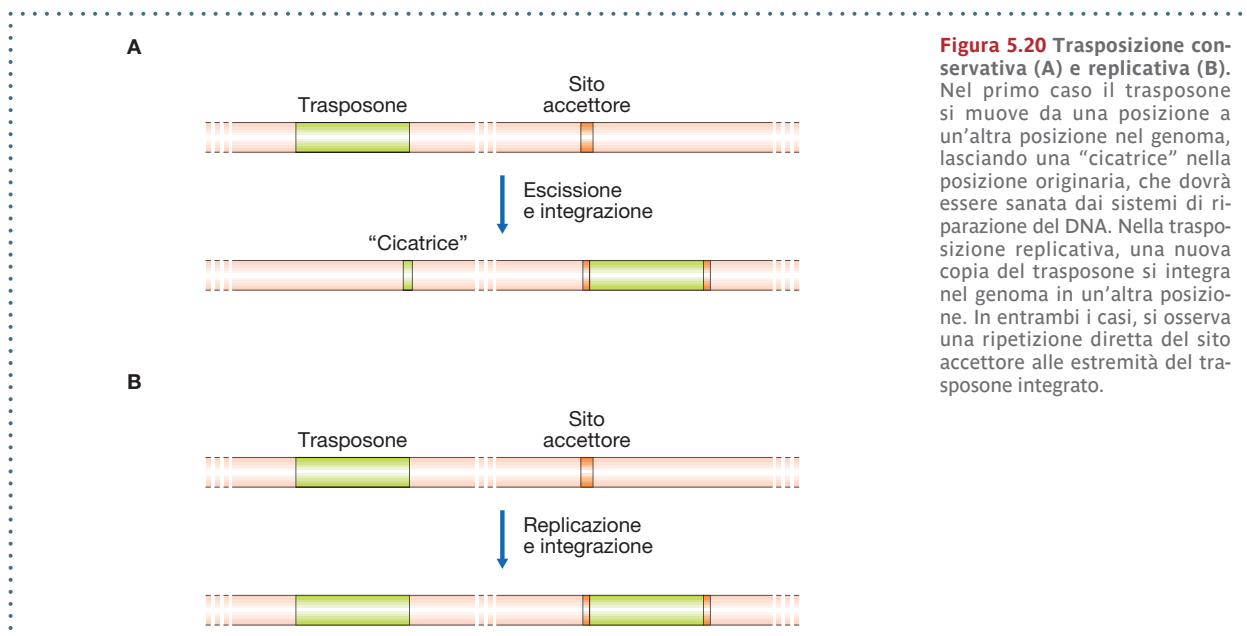


Figura 5.20 Trasposizione conservativa (A) e replicativa (B). Nel primo caso il trasposone si muove da una posizione a un'altra posizione nel genoma, lasciando una "cicatrice" nella posizione originaria, che dovrà essere sanata dai sistemi di riparazione del DNA. Nella trasposizione replicativa, una nuova copia del trasposone si integra nel genoma in un'altra posizione. In entrambi i casi, si osserva una ripetizione diretta del sito accettore alle estremità del trasposone integrato.

Tabella 5.6 Frequenza approssimativa dei diversi elementi trasponibili in vari organismi. La composizione e la percentuale relativa dei diversi elementi trasponibili presenti nel genoma variano enormemente da specie a specie.

Specie	Nome comune	Genoma (Mpb)	Numero di geni codificanti proteine	Retro-trasposoni LTR (%)	Retro-trasposoni LINE/SINE (%)	Trasposoni a DNA (%)	Genoma occupato da elementi trasponibili (%)
<i>Saccharomyces cerevisiae</i>	Lievito di birra	12	5800	100,0	0,0	0,0	3
<i>Caenorhabditis elegans</i>	Verme nematode	100	18400	1,7	6,9	91,4	6
<i>Arabidopsis thaliana</i>	Arabetta comune	125	25500	46,0	5,0	49,0	14
<i>Drosophila melanogaster</i>	Moscerino della frutta	180	13600	69,2	22,7	8,1	15
<i>Homo sapiens</i>	Uomo	3000	25000	18,6	75,1	6,3	44
<i>Oryza sativa</i>	Riso	400	37500	56,2	3,7	40,1	35
<i>Zea mays</i>	Mais	3200	50000	95,0	1,7	3,3	60

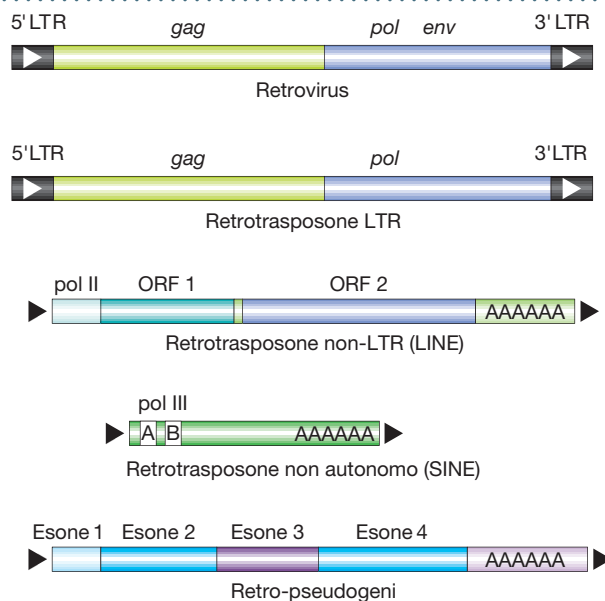


Figura 5.21 Struttura dei principali elementi retrotrasponibili. I retrovirus codificano per tre proteine (*gag*, antigene gruppo-specifico; *pol*, trascrittasi inversa; *env*, proteina di rivestimento). Il loro genoma è costituito da una molecola di RNA che prima dell'integrazione nel genoma si converte in una molecola di DNA circolare a doppio filamento. Gli elementi LTR sono simili ai retrovirus ma essendo privi del gene per la proteina *env* non sono in grado di for-

mare particelle virali. Gli elementi LINE contengono un insolito promotore interno per l'RNA polimerasi II e codificano per le proteine necessarie per la loro trasposizione. Gli elementi SINE utilizzano un promotore interno per l'RNA polimerasi III (box A e B) e utilizzano una trasposizione LINE-dipendente come pure i retro-pseudogeni. Nel sito di inserimento di questi elementi si osservano brevi tratti ripetuti indicati in figura dai piccoli triangoli neri alle estremità.

mità di dimensioni pari a 250-600 pb (LTR). Gli elementi LTR sono molto abbondanti in funghi, piante e insetti. Questi, per esempio, comprendono l'elemento *Ty* in lievito e l'elemento *copia* in *Drosophila*. Questi elementi sono privi del gene *env*, o ne posseggono una copia *non funzionante*, per cui non sono in grado di formare particelle virali trasmissibili. Gli elementi LTR dotati del gene *env* non funzionale sono anche denominati retrovirus endogeni. Per esempio, il genoma umano è stato colonizzato da migliaia di copie di HERV (*Human Endogenous RetroVirus*) originate da diverse famiglie di virus. Il gene della sincitina, specifico dei mammiferi, che gioca un ruolo cruciale nello sviluppo della placenta, deriva proprio da HERV. Quindi, come abbiamo già detto, alcuni elementi trasponibili possono essere addomesticati e contribuiscono in modo determinante al nostro genoma.

Gli **elementi LINE** (o trasposoni non-LTR) hanno una lunghezza di circa 6-7 kpb, contengono un promotore per l'RNA polimerasi II, una o due ORF e un segnale di poliadenilazione al 3'. L'ORF1 codifica per una proteina in grado di legare il proprio RNA, mentre l'ORF2 codifica per una proteina dotata sia dell'attività di trascrittasi inversa che di endonucleasi. Il meccanismo di retrotrasposizione comporta la formazione di un complesso ribonucleoproteico costituito dall'RNA LINE legato ai prodotti delle ORF1 e ORF2. Questo, una volta traslocato nel nucleo, effettua un taglio endonucleolitico a livello del sito di inserzione, che innesca il processo di retrotrascrizione a partire dall'estremità 3'. Il processo di retrotrascrizione risulta spesso incompleto con conseguente incorporazione di elementi LINE inattivi, tronchi al 5'. Il processo di incorporazione viene poi completato con la saldatura a livello dei tagli sfalsati prodotti dall'attività endonucleasica, che genera piccole sequenze ripetute alle estremità dell'elemento mobile integrato. Gli elementi LINE occupano oltre il 20% del genoma umano, ma solo un numero ridottissimo di copie (80-100) risulta attivo.

Gli **elementi SINE** sono elementi mobili non autonomi, di lunghezza compresa tra 100 e 400 pb. Hanno generalmente un promotore interno per l'RNA

polimerasi III e una regione ricca in A all'estremità 3'. La loro trasposizione richiede una trascrittasi inversa sintetizzata da altri retroelementi come i LINE.

Nel genoma dell'uomo e degli altri primati sono presenti tre diverse sottofamiglie di elementi SINE, l'elemento Alu ancora attivo e gli elementi inattivi MIR e Ther2/MIR3. L'elemento Alu, il più comune tra i primati, è lungo circa 300 pb, ed è presente nel genoma umano per oltre un milione di copie.

Gli elementi trasponibili giocano un ruolo fondamentale nell'evoluzione dei genomi. Essi, infatti, contribuiscono sia in modo attivo che passivo alla continua generazione di novità genetiche che saranno poi sottoposte al vaglio della selezione naturale. Il loro ruolo attivo consiste nell'integrazione in diversi punti del genoma, contribuendo a formare nuovi geni, o parti di essi, attraverso la loro parziale esonizzazione (**fig. 5.14**), oppure disseminando elementi di regolazione dell'espressione genica. Nella modalità passiva essi semplicemente favoriscono i processi ricombinativi alla base dei riarrangiamenti genomici (*vedi* cap. 8).

Le alterazioni strutturali dovute alla mobilitazione di elementi trasponibili possono avere effetti deleteri quando alterano l'espressione di geni critici. Infatti, disfunzioni geniche dovute all'inserimento in specifiche posizioni nel genoma di elementi trasponibili sono responsabili di numerose patologie, incluse diverse tipologie di cancro. Tuttavia, come detto sopra, l'attività degli elementi trasponibili può anche avere effetti positivi favorendo l'introduzione di variabilità genetica che facilita l'adattamento di popolazioni e specie alle più disparate forme di stress.

Duplicazioni segmentali

Le duplicazioni segmentali contribuiscono alla componente ripetitiva del genoma. Queste sono ripetizioni di segmenti genomici comunemente osservate in genomi di animali e piante, non riconducibili a elementi trasponibili, di lunghezza $\geq 10\,000$ pb e identità $> 90\%$. Le duplicazioni segmentali possono essere **intracromosomiche**, tipicamente in tandem, o **intercromosomiche**. Si ipotizza che queste ultime possano essere state prodotte da una duplicazione in tandem poi seguita da traslocazione.

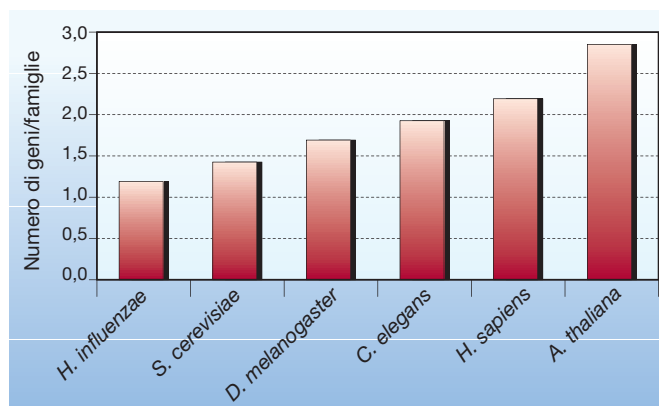
Si possono osservare ripetizioni segmentali anche molto estese, come nel cromosoma Y umano che presenta un blocco ripetuto di 1,45 Mpb. Sono localizzate in prevalenza nelle regioni adiacenti ai centromeri e raramente presenti nelle regioni più distali di ciascun braccio del cromosoma. Le duplicazioni sono un'importante fonte di variabilità genetica tra individui della popolazione umana e, dato che possono estendersi su più geni, comportano la variazione del numero di copie di determinati geni, tra diversi individui. La variazione del numero di copie di uno o più geni (*Copy Number Variation*, CNV) comporta variazioni del livello di espressione genica e può essere responsabile di patologie come il cancro della mammella, l'autismo e alcune malattie autoimmuni. Inoltre, queste regioni sono soggette a riarrangiamenti cromosomici responsabili di malattie genetiche come le sindromi DiGeorge e Charcot-Marie-Tooth. Uno studio condotto sul genoma umano analizzando 250 individui di 4 diverse popolazioni ha individuato circa 1500 loci CNV.

Famiglie geniche

I geni che codificano per proteine fanno parte della frazione non ripetuta del DNA genomico eucariotico. In realtà, la maggior parte dei geni non è presente in singola copia. Infatti, abbastanza frequentemente, si trovano due o più copie correlate dello stesso gene, il cui insieme viene chiamato **famiglia genica**, a indicare il fatto che si tratta di copie provenienti, per duplicazione, da uno stesso gene ancestrale.

Pertanto, il numero di geni distinti all'interno di un genoma corrisponde al numero di geni unici (che potrebbero essere considerati famiglie geniche costituite da un unico membro) più il numero di famiglie geniche costituite da due o più membri. Alcune famiglie geniche possono essere anche molto numerose, come quella per i recettori dell'olfatto, che comprende circa un migliaio di geni nei mammiferi.

Figura 5.22 Rapporto tra il numero dei geni contenuto nel genoma e il numero di famiglie geniche.



Se consideriamo il numero di famiglie geniche, le differenze relative al numero di geni nei genomi eucariotici illustrate in **figura 5.8** si riducono ulteriormente.

In **figura 5.22** si vede come la dimensione media delle famiglie geniche aumenti passando dai procarioti agli eucarioti unicellulari (lievito), e da questi agli eucarioti multicellulari. Il fatto che i genomi delle piante contengano un numero di geni maggiore di quello dell'uomo e degli altri mammiferi si spiega per il fatto che le prime contengono un maggior numero di geni duplicati.

Le duplicazioni geniche che portano all'espansione delle famiglie geniche possono aver luogo attraverso diversi meccanismi: (1) poliploidizzazione del genoma; (2) duplicazione di estese regioni genomiche; (3) duplicazione di singoli geni.

Nel primo caso si ha la duplicazione dell'intero genoma, e quindi dell'intero campionario dei geni, generalmente seguita dalla perdita di una frazione rilevante dei geni duplicati. L'analisi evolutiva dei genomi ha permesso di identificare diversi di questi eventi. Per esempio, è noto che nel corso dell'evoluzione dagli invertebrati ai vertebrati sono avvenuti due eventi di completa duplicazione del genoma, come testimoniato dall'osservazione che il cluster dei geni omeotici (*Hox*) è quadruplicato nei mammiferi rispetto a *Drosophila* (**fig. 5.23**). È quindi probabile che l'evoluzione di questa famiglia genica abbia avuto luogo attraverso un processo a due stadi, ovvero da un evento di duplicazione di un gene *Hox* primordiale che ha originato il cluster *Hox* negli invertebrati, seguito da eventi di duplicazione dell'intero cluster fino a produrre i cluster multipli che oggi osserviamo nell'uomo e negli altri vertebrati.

Lo studio di numero, organizzazione genomica e sequenze di alcune famiglie geniche in specie diverse più o meno correlate tra loro ha permesso di far luce su importanti aspetti dell'evoluzione. Le famiglie geniche studiate sono moltissime, il caso emblematico che in genere viene portato come esempio è la famiglia di geni delle globine, le proteine che legate in forma tetramerica all'eme costituiscono l'emoglobina dei globuli rossi del sangue dei vertebrati.

Tutti i geni delle globine derivano da un singolo gene ancestrale che circa 800 milioni di anni fa (**fig. 5.24**) ha generato per duplicazione i geni delle mioglobine e del progenitore delle α -globine e delle β -globine. Quest'ultimo ha poi generato, in seguito a una duplicazione avvenuta circa 500 milioni di anni fa, i geni delle globine α e β , che hanno subito ulteriori eventi di duplicazione nelle diverse linee di discendenza. Una globina funzionale è costituita da due subunità α e due subunità β .

Nel genoma umano si osserva il cluster delle α -globine nel cromosoma 16 e quello delle β -globine nel cromosoma 11, costituiti rispettivamente da tre e cinque geni funzionali, oltre a diversi pseudogeni. I prodotti di questi geni hanno specializzato la loro funzione in modo da conferire una maggiore affinità per l'ossigeno alle forme embrionali ($\alpha_2\epsilon_2$, $\zeta_2\epsilon_2$, $\zeta_2\gamma_2$) e fetali ($\alpha_2\gamma_2$) rispetto alle forme

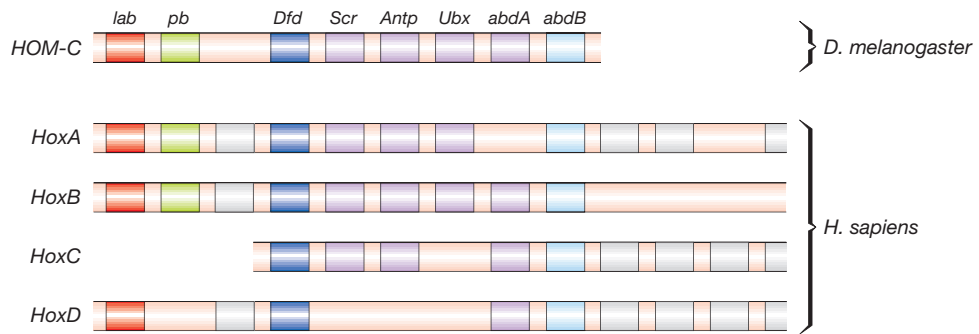


Figura 5.23 Famiglia genica dei geni omeotici in *Drosophila melanogaster* (HOM-C) e nell'uomo (Hox). Il cluster dei geni omeotici (Hox), quadruplicato nei vertebrati rispetto agli invertebrati, è

conservato sia nella composizione che nella disposizione dei geni (i geni omologhi sono nella stessa colorazione).

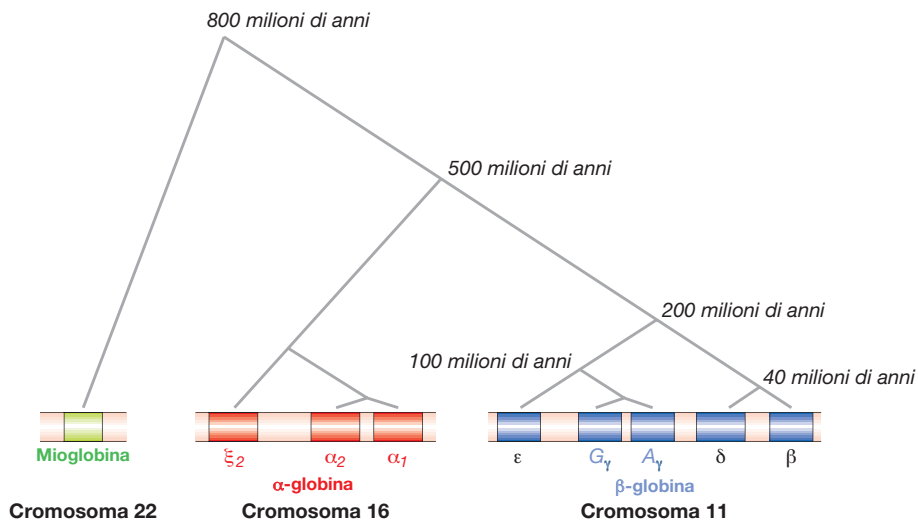


Figura 5.24 Evoluzione della famiglia genica delle globine umane.

adulte ($\alpha_2\beta_2$, $\alpha_2\delta_2$). Questa caratteristica è necessaria per lo scambio dell'ossigeno con il sangue materno durante la gestazione.

Come abbiamo visto per le globine, sono noti molti altri casi in cui i diversi membri di una stessa famiglia genica possono formare diversi complessi eteromericici, il cui assortimento e composizione ne determina l'attività funzionale.

Lo studio delle relazioni evolutive tra i geni di una famiglia genica, in termini di relazioni filogenetiche e analisi della velocità di sostituzione nucleotidica o amminoacidica (vedi par. 21.8), è molto importante per comprendere la storia evolutiva e la specializzazione funzionale dei diversi membri di una famiglia genica. Difatti, a seguito del processo di duplicazione genica alla base del processo di espansione del numero di membri di una famiglia genica, una delle copie rimane, di norma, molto simile alla copia originaria preservandone la funzione mentre l'altra, ridondante e quindi meno soggetta a costrizioni funzionali, può accumulare mutazioni che portano all'acquisizione di una nuova funzione.

Geni reiterati ed evoluzione parallela

Nel caso di geni che debbano essere espressi a livelli elevati nella cellula, la duplicazione genica risulta vantaggiosa per il fatto che la presenza di copie multiple del gene garantisce un proporzionale aumento del livello di espressione.

Questa è la ragione per cui i geni per rRNA, tRNA e alcune proteine come gli istoni sono presenti in copie multiple nel genoma. Negli eucarioti questi geni ripetuti, detti anche **geni reiterati**, sono organizzati in tandem, cioè con un arrangiamento testa-coda, e costituiscono dei gruppi alle volte anche molto numerosi. Descriviamo qui, come esempio di organizzazione ed evoluzione dei geni reiterati, il caso del gene per il precursore degli rRNA (*vedi par. 12.1*). Questo gene è presente in copie multiple in tutti gli organismi. Negli eucarioti le copie sono organizzate in tandem in un locus cromosomico chiamato **organizzatore nucleolare** o **NOR** (*Nucleolar Organizer Region*) (*vedi finestra 12.2*). In alcune specie, come lievito, *Drosophila* e *Xenopus*, tutte le copie del gene per l'rRNA sono raggruppate in un unico locus NOR, in altre, come l'uomo e in genere i mammiferi, sono ripartite in più loci NOR. La figura 12.2 mostra un tratto del genoma umano contenente alcune delle copie reiterate del gene per l'rRNA, in cui si può osservare l'alternanza tra regioni trascritte, i geni veri e propri, e spaziatori non trascritti (NTS, *Non Transcribed Spacer*).

È stato osservato che i geni per l'rRNA, anche se evolutivamente conservati, presentano alcune differenze tra le diverse specie, indicando che possono subire mutazioni e andare incontro a cambiamenti evolutivi. È stato d'altra parte osservato che, in ciascuna specie, le copie reiterate del gene sono sostanzialmente identiche tra loro in modo da produrre un rRNA omogeneo che andrà a costituire ribosomi tutti uguali tra loro. Queste due osservazioni sono contraddittorie e mettono in evidenza un paradosso: infatti le mutazioni, essendo eventi casuali, dovrebbero col passare del tempo causare una progressiva divergenza delle copie geniche su cui, essendo numerose, la selezione non ha possibilità di azione. Per spiegare questo paradosso è stato proposto che i geni per l'rRNA, e anche gli altri geni reiterati, vadano incontro a una "evoluzione parallela" (o "concertata" o "coincidente") delle copie geniche, un meccanismo cioè che al tempo stesso permetta l'occorrenza di mutazioni mantenendo però tutte le copie uguali tra loro. Qual è il meccanismo alla base di questa evoluzione parallela? Su questo ancora non è stata fatta completa chiarezza. Sono stati proposti vari modelli che si basano sostanzialmente su tre meccanismi genetici già noti in altri contesti.

1. **Crossing over diseguale.** Quando nel genoma esistono due o più copie di una sequenza adiacenti o, comunque, non molto distanziate, queste possono andare incontro a crossing over diseguale, cioè, come mostrato in **figura 5.25**, a un crossing over tra due copie non corrispondenti. Il crossing over diseguale è un processo sicuramente esistente ed è implicato nell'evoluzione delle famiglie geniche per duplicazione e divergenza. Se applicato a un gruppo di sequenze reiterate come i geni per l'rRNA, il crossing over diseguale fa sì che una cellula figlia riceva una o più copie geniche in meno, e l'altra una o più copie in più. Come si può osservare nello schema della figura, la continua riduzione ed espansione del numero di copie mediante crossing over diseguali può portare, dopo molti eventi di ricombinazione, alla perdita di alcuni tipi di copie e all'aumento numerico di altre. Questo tenderebbe a omogenizzare le copie in ciascun gruppo di geni e faciliterebbe l'azione della selezione.
2. **Conversione genica.** Un altro noto processo genetico che potrebbe spiegare l'evoluzione concertata dei geni reiterati è la "conversione genica", un processo ricombinativo mediante il quale una di due sequenze omologhe ma non identiche viene corretta sull'altra. Il meccanismo della conversione genica è descritto nella finestra 8.1, dove la vediamo in azione tra due alleli ed è responsabile di alcuni casi di segregazione non mendeliana, e la ritroveremo anche nel fenomeno del cambiamento di tipo sessuale in lievito, dove è responsabile della conversione del locus attivo MAT su uno dei due loci, HML o HMR, sullo stesso cromosoma (*vedi fig. 8.15*). Nel contesto di cui stiamo parlando ora, un meccanismo analogo alla conversione genica potrebbe portare alla perdita di una copia mutata per correzione su una copia normale, oppure all'espansione

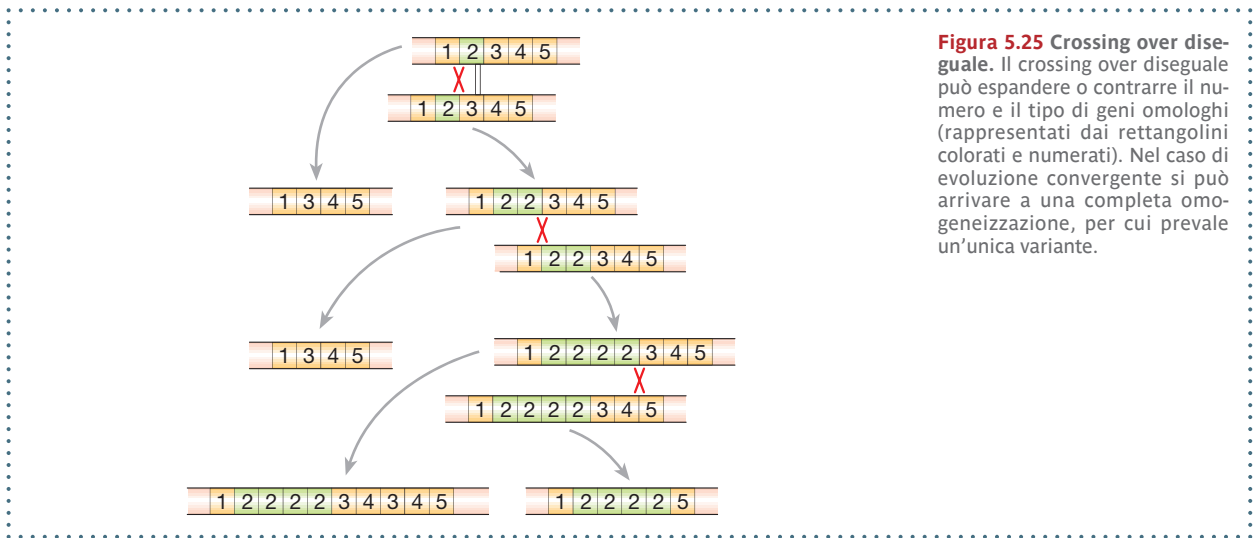


Figura 5.25 Crossing over diseguale. Il crossing over diseguale può espandere o contrarre il numero e il tipo di geni omologhi (rappresentati dai rettangolini colorati e numerati). Nel caso di evoluzione convergente si può arrivare a una completa omogeneizzazione, per cui prevale un'unica variante.

della copia mutata per correzione su di essa delle copie non mutate. Essendo le copie dei geni reiterate anche molte decine o centinaia, il meccanismo di conversione dovrebbe reiterarsi lungo il gruppo di copie geniche.

3. **Perdita e sostituzione delle copie reiterate.** Un terzo possibile meccanismo che può spiegare l'evoluzione parallela delle copie di geni ripetuti ipotizza che, in particolari occasioni, gran parte o l'intero gruppo di copie geniche vada perso, per esempio per escissione conseguente a ricombinazione intramolecolare, e venga sostituito da un nuovo gruppo derivato per replicazione da una o poche copie. Nel caso dei geni per l'rRNA, un processo di questo genere potrebbe avvenire durante l'oogenesi, che vede in molte specie la presenza di copie amplificate extracromosomiche di questi geni.

La cosa più probabile è che tutti questi meccanismi, e possibilmente altri, concorrano a mantenere l'omogeneità delle copie reiterate e permettano un'evoluzione concertata dei geni reiterate.

5.4 Genoma mitocondriale

La maggior parte delle cellule eucariotiche contiene i **mitocondri**, organelli citoplasmatici dotati di un sistema genetico indipendente, completo dell'apparato necessario per la sua espressione.

Vi sono numerose evidenze dell'**origine endosimbiontica** dei mitocondri a partire da un alfa-proteobatterio. È anche probabile che l'origine dei mitocondri coincida con la formazione delle prime cellule eucariotiche (fig. 5.1). Difatti, anche gli organismi eucariotici unicellulari, privi di mitocondri, denominati Archeozoi, posseggono delle vestigia tipicamente mitocondriali, lasciando supporre una perdita secondaria dei mitocondri.

Una delle ipotesi più accreditate sull'origine della cellula eucariotica, avvenuta oltre un miliardo di anni fa, suggerisce l'endosimbiosi tra un archeobatterio metanogeno consumatore di H_2 e un alfa-proteobatterio produttore di H_2 . L'alfa-proteobatterio avrebbe poi dato origine al mitocondrio trasferendo gran parte dei suoi geni nel genoma archeobatterico, successivamente compartimentalizzato nella struttura nucleare della primitiva cellula eucariotica. Secondo questa ipotesi l'origine del mitocondrio coincide con l'origine della primitiva cellula eucariotica. Pertanto, gli eucarioti monocellulari privi di mitocondri (per es. *Giardia lamblia*, *Trichomonas vaginalis*) avrebbero subito una perdita secondaria dei mitocondri. A supporto di questa ipotesi endosimbiontica vi è l'osservazione della

Tabella 5.7 Caratteristiche divergenti dei genomi mitocondriali di animali, funghi e piante.

	Animali	Funghi	Piante
Dimensioni	14-42 kpb	17-180 kpb	180-2400 kpb
Contenuto di DNA non codificante	Basso	Variabile	Elevato
Tasso di mutazione	Alto	Basso	Molto basso
Ricombinazione	No	Sì/No	Sì
Presenza di introni	No	Sì/No	Sì
Codice genetico universale	No	Sì/No	Sì

natura chimerica del genoma nucleare, dove una parte dei geni risulta più simile alla controparte archeobatterica e l'altra alla controparte eubatterica.

Nel corso dell'evoluzione il genoma dell'originario organismo endosimbionte si è andato progressivamente contraendo con il concomitante trasferimento di geni verso il genoma nucleare. Tuttavia, le dinamiche evolutive che hanno portato ai genomi mitocondriali che oggi osserviamo negli eucarioti unicellulari e multicellulari hanno subito percorsi piuttosto divergenti. Per questa ragione si osservano differenze significative tra diversi organismi per molte caratteristiche, come le dimensioni, il contenuto di DNA non codificante, il tasso di mutazione, la presenza di introni, il codice genetico (**tab. 5.7**). Le caratteristiche più primitive dei genomi mitocondriali si ritrovano attualmente in alcuni protisti (ad es. *Reclinomonas americana*).

Per quanto riguarda la forma, i genomi mitocondriali sono generalmente costituiti da una singola molecola di DNA circolare, ma vi sono anche casi di genomi costituiti da una o più molecole lineari, oppure da più molecole circolari.

Il contenuto in geni dei genomi mitocondriali è sostanzialmente conservato, comprendendo essenzialmente geni coinvolti nel metabolismo respiratorio e nella traduzione. Infatti, si osserva invariabilmente la presenza di due o tre geni per rRNA, di un numero variabile di geni per tRNA e di alcuni geni che codificano per subunità del sistema della fosforilazione ossidativa (OxPhos).

Nella maggior parte dei gruppi tassonomici, sia piante che animali, alcuni geni mitocondriali contengono introni in grado di operare autosplicing, sia della classe I che della classe II (*vedi par. 13.1*). Tra le eccezioni ci sono i cordati, inclusa la specie umana, i cui geni mitocondriali invece non contengono introni.

Il **DNA mitocondriale (mtDNA)** umano è una molecola circolare costituita da 16569 pb che contiene 37 geni di cui 13 codificano per proteine del sistema OxPhos, 22 tRNA e 2 rRNA (**fig. 5.26**). Le 13 proteine codificate dal mtDNA entrano nella costituzione di 3 dei 4 complessi respiratori e dell'ATP sintasi. La restante parte delle circa 80 proteine del sistema OxPhos, al pari di tutte le proteine che partecipano alla formazione e assemblaggio dei complessi stessi, sono invece codificate da geni nucleari. In particolare, il genoma mitocondriale codifica per:

- sette subunità del complesso I (NADH deidrogenasi): ND1, ND2, ND3, ND4, ND4L, ND5, ND6;
- una subunità del complesso III (citocromo c-coenzima Q ossidoreduttasi), l'apocitocromo b;
- tre subunità del complesso IV (citocromo ossidasi): COX1, COX2 e COX3;
- due subunità del complesso V (ATP sintasi): ATPasi 6 e ATPasi 8.

L'organizzazione del genoma mitocondriale nell'uomo e negli altri vertebrati è molto diversa da quella nucleare. In primo luogo è molto compatta, non essendoci né introni né consistenti regioni intergeniche (a eccezione delle due regioni non codificanti a funzione regolatrice gli unici spaziatori intergenici presenti hanno lunghezza inferiore alle 8 pb). Addirittura alcuni geni hanno le regioni codificanti parzialmente sovrapposte. In secondo luogo, come mostrato in **figura 5.26**, i geni

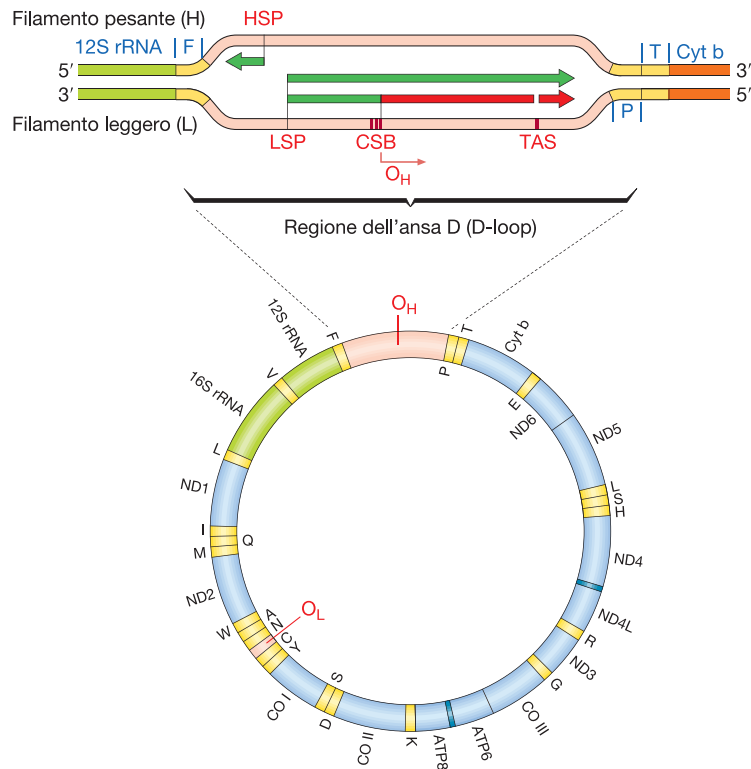


Figura 5.26 Struttura del genoma mitocondriale dei metazoi. Il genoma mitocondriale dei metazoi, costituito da un DNA circolare a doppio filamento, contiene generalmente 37 geni distribuiti sui due filamenti H e L (le etichette dei geni corrispondenti sono rispettivamente all'esterno e all'interno del circolo) che comprendono 22 geni per tRNA (in giallo) indicati dal codice a una lettera del corrispondente amminoacido, 2 geni per rRNA (in verde) e 13 geni codificanti proteine (in blu). Le regioni in blu più scuro corrispondono alle regioni di sovrapposizione tra ATP6 e ATP8, e ND4

e ND4L. Sono anche indicate le origini di replicazione per il filamento pesante (O_H), all'interno della principale regione non codificante denominata D-loop, e leggero (O_L). Nell'espansione della regione del D-loop mostrata in alto sono evidenziati i promotori dell'elica leggera (LSP) e pesante (HSP). A partire dal promotore LSP viene anche generato l'rRNA che funge da innesco (in verde) per la replicazione del DNA mitocondriale (in rosso), che normalmente resta quiescente a livello delle sequenze TAS. Le sequenze conservate CSB e TAS sono rappresentate da rettangolini rossi.

mitocondriali sono distribuiti in modo asimmetrico tra i due filamenti del DNA, con un filamento codificante per soli 9 geni (8 tRNA e una proteina) e l'altro filamento codificante per i restanti 28 geni. In terzo luogo, mentre i geni nucleari sono trascritti come mRNA monocistronici, i geni mitocondriali sono trascritti in unità policistroniche che partono da almeno due promotori situati all'interno dell'unica ampia regione non codificante denominata *Displacement-loop* (**D-loop**): *Light Strand Promoter* (**LSP**) e *Heavy Strand Promoter* (**HSP**). Poiché i due filamenti del mtDNA presentano asimmetria nella composizione in basi rilevabile mediante gradiente denaturante di cloruro di cesio, essi sono stati denominati l'uno filamento pesante o *heavy* (H) e l'altro filamento leggero o *light* (L). A partire da HSP, situato sul filamento *heavy*, sono trascritti come un unico RNA precursore un totale di 28 geni mitocondriali. I restanti 9 geni, cioè il gene per la subunità 6 della NADH deidrogenasi e otto geni per tRNA, sono invece trascritti a partire dal promotore LSP, ancora una volta in forma di un precursore policistronico.

Le uniche due sequenze non codificanti dell'intero genoma mitocondriale umano sono sede di importanti funzioni regolatrici. Infatti la regione del D-loop, di circa 1,1 Kb, contiene, oltre ai promotori HSP e LSP, l'origine di replicazione del filamento pesante (O_H) e le sequenze conservate conosciute come **CSB** (*Conserved Sequences Blocks*) e **TAS** (*Termination-Associated Sequences*). I CSB sono i siti dove avviene la maturazione dei trascritti iniziati da LSP e utilizzati come inneschi per la sintesi del filamento H di DNA nascente, che può essere terminata prematura-

mente in corrispondenza dei siti TASs generando il cosiddetto 7S DNA D-loop (500-700 pb) della tripla elica.

Infatti, come mostrato nella parte alta di **figura 5.26**, in seguito all'inizio della replicazione si possono verificare due diversi eventi: la sintesi del filamento 7S DNA, evento più comune, oppure, in un minor numero di casi, la replicazione oltrepassa le sequenze TAS e prosegue per tutta la lunghezza del genoma. Il meccanismo che stabilisce il procedere o meno dell'allungamento è ancora sconosciuto. Il filamento nascente durante l'allungamento separa i filamenti della molecola madre fino a raggiungere l' O_L dalla quale prende inizio la sintesi del filamento L in direzione antioraria.

L'origine di replicazione del filamento leggero (O_L) è contenuta all'interno della seconda regione non codificante di soli 30 nucleotidi, posizionata a circa 2/3 della molecola di mtDNA rispetto alla O_H in un cluster di 5 geni per tRNA noto come WANCY (dai codici a una lettera degli amminoacidi trasportati dai tRNA codificati). Pertanto, la replicazione della molecola di mtDNA risulta essere asincrona e asimmetrica (con il filamento H, la cui sintesi inizia precocemente rispetto al filamento L), benché ultimamente meccanismi di replicazione simmetrica siano stati accertati. Secondo tale modello la sintesi dei nuovi filamenti avviene mediante due modalità distinte e dipende dalle condizioni cellulari.

Indipendentemente dal modello di replicazione del mtDNA, il macchinario replicativo o replisoma necessario è costituito almeno da due enzimi chiave per la sintesi del DNA: una polimerasi mitocondriale o polimerasi γ (*POLG*) e un'elicasi (*Twinkle*).

All'interno della matrice mitocondriale le molecole di mtDNA sono associate in strutture simili al nucleoide batterico, composte da diverse molecole di mtDNA e proteine, associate alla membrana mitocondriale interna. Tra queste vi sono il fattore di trascrizione del DNA mitocondriale (mtTFA: *mitochondrial Transcription Factor A*), una proteina che lega il DNA a singola elica (mtSSB: *Single-Strand DNA Binding*) più altre proteine e fattori. La distribuzione dei nucleoidi, durante gli eventi di fusione e fissione e nella citochinesi, è un evento importante per la segregazione del genoma mitocondriale.

I tRNA mitocondriali sono più piccoli dei tRNA citosolici, codificati nel nucleo, e possono avere strutture secondarie insolite, come la completa assenza del braccio D. Essi hanno un vacillamento più ampio dato che l'interazione codone-anticodone è sostanzialmente basata su due soli nucleotidi, in quanto la U nella prima posizione dell'anticodone riconosce tutti e quattro i nucleotidi nella terza posizione del codone. Nei vertebrati, oltre ad agire durante la traduzione, i tRNA mitocondriali hanno inoltre il ruolo di elementi di punteggiatura per guidare la maturazione dell'RNA precursore. Infatti i tagli dell'RNA precursore policonico per liberare rRNA, mRNA e tRNA mitocondriali avvengono proprio a livello delle due estremità delle strutture a trifoglio dei tRNA.

La struttura e il contenuto genico del genoma mitocondriale umano sono rimasti sostanzialmente conservati nei vertebrati.

In molte specie appartenenti a diversi gruppi tassonomici il genoma mitocondriale adotta un codice genetico specifico che presenta alcune deviazioni rispetto al codice genetico universale. Solo i mitocondri dei protisti e delle piante adottano il codice genetico universale. Come abbiamo già descritto nel capitolo 3 (*vedi par. 3.6*) le deviazioni dal codice genetico universale osservate nel genoma mitocondriale dei vertebrati sono AGA/AGG: Arg $\rightarrow$ Stop; AUA: Ile $\rightarrow$ Met; UGA: Stop $\rightarrow$ Trp. Le deviazioni rispetto al codice universale pongono una barriera al trasferimento di geni dal genoma mitocondriale a quello nucleare.

Vi sono più molecole di DNA mitocondriale per mitocondrio (5-10) e molti mitocondri per cellula. Il numero di mitocondri per cellula è correlato sia alle dimensioni della cellula stessa sia al suo fabbisogno energetico. Per esempio, nell'uomo, gli oociti contengono circa 200 000 mitocondri, mentre le cellule epatiche circa 1000.

FINESTRA 5.2

PATOLOGIE MITOCONDRIALI

I difetti del metabolismo mitocondriale si manifestano con un ampio spettro di patologie umane. Dal punto di vista epidemiologico studi recenti hanno confermato che le malattie mitocondriali, a lungo considerate di puro interesse accademico, sono tra i più comuni disordini genetici. Infatti, studi condotti sulla popolazione pediatrica e adulta hanno evidenziato un'incidenza di circa 1:8000 nelle popolazioni caucasiche, ponendo tali malattie tra le più frequenti patologie geneticamente determinate. I dati epidemiologici sono probabilmente una sottostima della situazione reale a causa della variabilità fenotipica di queste patologie, motivo di difficoltà per il clinico nel diagnosticarle. Anche alcune patologie neurodegenerative più comuni, tra cui forme familiari del morbo di Parkinson, trovano la loro causa in alcuni geni per proteine mitocondriali. I difetti della funzionalità del **sistema della fosforilazione ossidativa (OxPhos)** si ripercuotono sui tessuti maggiormente dipendenti dal metabolismo aerobio come il sistema nervoso centrale (SNC), il muscolo scheletrico e cardiaco, l'occhio, il fegato, il pancreas e l'organo del Corti nell'orecchio interno. Disfunzioni del sistema OxPhos determinano l'aumento dello stato di riduzione della cellula che sbilanciano l'equilibrio piruvato/lattato verso l'accumulo di lattato nel sangue, nelle urine e nel liquido cerebrospinale. Inoltre disfunzioni OxPhos possono sensibilizzare la cellula all'induzione della via mitocondriale dell'apoptosi. Poiché i mitocondri compiono svariate funzioni in differenti tessuti, le malattie mitocondriali sono numerose e succede che identiche mutazioni del mtDNA possano essere associate a malattie differenti. È vero anche l'inverso: differenti mutazioni nel mtDNA e nDNA possono portare alla stessa manifestazione clinica. Una classificazione generale, su basi genetiche, distingue due grandi gruppi:

1. patologie dovute a mutazioni del mtDNA;
2. patologie dovute a mutazioni in geni nucleari codificanti proteine coinvolte nel metabolismo mitocondriale.

Mitocondriopatie causate da mutazioni del DNA mitocondriale

Il genoma mitocondriale, che si è dimostrato essere responsabile di un'ampia gamma di sindromi cliniche, può essere colpito da numerose mutazioni. Queste mutazioni sono disseminate lungo l'intera molecola, tuttavia sono emersi alcuni "hot spot" (m.3243G>A, m.8344G>A, m.8993T>G/C, m.11778A>G; la posizione si riferisce alla sequenza di riferimento NC_012920) oltre ai riarrangiamenti su larga scala (delezioni o duplicazioni parziali), che interessano regioni che comprendono diversi geni mitocondriali e sono per lo più eteroplasmici. Le mutazioni puntiformi possono essere eteroplasmiche od omoplasmiche e interessare sia geni coinvolti nella sintesi proteica (tRNA, rRNA), che geni codificanti per proteine strutturali dei complessi del sistema OxPhos.

Mutazioni puntiformi del DNA mitocondriale

La distinzione tra mutazioni patogeniche e i numerosi polimorfismi non può essere immediata a causa dell'elevata fre-

quenza mutazionale del mtDNA. Le mutazioni patogeniche sono solitamente caratterizzate da elevata conservazione filogenetica sia a livello amminoacidico che nucleotidico (tRNA e rRNA), alterazione funzionale del prodotto genico, segregazione con il fenotipo patologico, correlazione tra la percentuale di molecole mutate e la gravità del fenotipo clinico, identificazione delle mutazioni in famiglie affette provenienti da popolazioni umane etnicamente distinte.

Le mutazioni più frequenti riscontrate allo stato eteroplasmico sono:

- m.3243G>A, associata all'encefalomiopatia mitocondriale con acidosi lattica ed episodi di tipo stroke (MELAS);
- m.8344G>A, associata all'epilessia mioclonica con RRF (MERRF);
- m.8993T>G e m.8993T>C, associate alla neuropatia con atassia e retinite pigmentosa (NARP) e alla MILS (sindrome di Leigh ereditata per via materna).

Le più comuni mutazioni riscontrate allo stato omoplasmico comprendono le mutazioni m.11778A>G, m.3460A>G e m.14484T>C, associate alla neuropatia ottica ereditaria di Leber (LHON).

Grandi riarrangiamenti del DNA mitocondriale

I riarrangiamenti su larga scala del mtDNA comprendono le **delezioni singole e multiple** e le **duplicazioni parziali**. Spesso però, le delezioni multiple del mtDNA derivano da un difetto nella comunicazione intergenomica (*vedi* oltre), in quanto sono associate a mutazioni in geni nucleari che intervengono nella replicazione e nel mantenimento del mtDNA. Le duplicazioni e le delezioni del mtDNA generalmente non sono trasmesse matrilinearmente come le mutazioni puntiformi, ma sono sporadiche e probabilmente si presentano durante l'oogenesi o precocemente nell'embriogenesi. Nei mitocondri contenenti le molecole delete di mtDNA il sistema OxPhos è difettivo, sia a causa della perdita dei geni che codificano i prodotti coinvolti nella sintesi proteica (tRNA e rRNA), sia a causa dell'eventuale perdita di geni che codificano per componenti strutturali dei complessi respiratori. I principali fenotipi clinici associati a riarrangiamenti singoli su larga scala del mtDNA sono tre: la sindrome di Kearns-Sayre, l'oftalmoplegia esterna progressiva e la sindrome di Pearson. La **sindrome di Kearns-Sayre (KSS)** è caratterizzata da oftalmoparesi, blocchi di branca cardiaca e retinite pigmentosa che si manifestano prima dei venti anni. I pazienti affetti da KSS mostrano tutti *ragged red fibers* (RRF) nell'istochimica del muscolo scheletrico. La patogenesi è legata generalmente a una singola delezione di circa 5 kb nella molecola del mtDNA (la cosiddetta *common deletion*). La **sindrome di Pearson (PS)** si manifesta nella prima infanzia ed è caratterizzata da pancitopenia sideroblastica congenita e, meno frequentemente, da una severa insufficienza del pancreas esocrino con conseguente malassorbimento. Nei soggetti che sopravvivono, solitamente si ha un'attenuazione dell'anemia e lo sviluppo di un quadro clinico tipico della KSS. Una singola delezione di circa 5 kb

(continua)

(*common deletion*) è la causa più comune anche dell'**oftalmoplegia esterna progressiva (PEO)**. Generalmente nella PEO le delezioni si possono individuare quasi esclusivamente nel muscolo, mentre nella KSS si possono trovare sia nel muscolo che nel sangue, anche se in quantità inferiore. Nella sindrome di Pearson, invece, la quantità di specie delete di mtDNA è paragonabile nel muscolo e nel sangue.

Mitocondriopatie dovute a difetti del DNA nucleare

Le mutazioni nei geni nucleari codificanti per proteine che svolgono la propria funzione nei mitocondri, circa 1500, sono causa di patologie mitocondriali a ereditarietà mendeliana e possono influenzare direttamente o indirettamente il funzionamento della respirazione mitocondriale. Si distinguono tre gruppi principali:

1. mutazioni in geni che codificano le subunità strutturali dei complessi della catena respiratoria. Il complesso I è il più colpito da mutazioni essendo il più grande. Mutazioni in questi geni determinano prevalentemente disordini neurologici progressivi con acidosi lattica in età pediatrica e, più spesso, sindrome di Leigh;
2. mutazioni in geni che codificano per proteine ausiliarie, indispensabili per l'assemblaggio, la formazione e la regolazione della biogenesi dei complessi;
3. mutazioni in geni che controllano il mantenimento e l'espressione del mtDNA. Alterazioni della stabilità e del numero di copie del mtDNA (comunicazione intergenomica) provocano disordini a trasmissione mendeliana in cui le mutazioni in geni nucleari causano alterazioni qualitative (delezioni multiple) e quantitative (deplezioni). Mutazioni in *POLG* e *POLG2*, che codificano per la subunità catalitica e accessoria della polimerasi γ , sono tra le cause più comuni.

A questo gruppo appartengono anche geni che controllano il pool dei nucleotidi della cellula. I geni che regolano l'espressione del mtDNA comprendono, per esempio, i fattori di traduzione e le amminoacil-tRNA sintetasi.

Al momento non è possibile elencare tutti i geni coinvolti, dato che quasi quotidianamente sono identificati nuovi geni che governano il metabolismo mitocondriale nei diversi tessuti. Vanno menzionati anche i geni per le proteine della plasticità e del movimento degli organelli e della composizione di membrana.

Diagnosi e terapia di patologie mitocondriali

Data la complessità della biochimica e della genetica mitocon-

driale le manifestazioni cliniche delle patologie mitocondriali sono estremamente eterogenee. I pazienti adulti, di solito, mostrano miopatia associata a coinvolgimento variabile del SNC: atassia, perdita dell'udito, neuropatie, retinopatia pigmentaria e, più raramente, disordini del movimento. Alcuni pazienti lamentano solo debolezza muscolare e intolleranza all'esercizio fisico. A causa della variabilità fenotipica la formulazione di una diagnosi per tali patologie è spesso difficoltosa e il coinvolgimento mitocondriale viene considerato solo dopo aver escluso altre possibilità diagnostiche. Una dettagliata anamnesi del paziente e l'esame clinico risultano fondamentali nell'iniziare un'indagine in un soggetto con sospetta patologia mitocondriale. Talvolta le indagini preliminari possono suggerire una specifica sindrome mitocondriale (MELAS, MERRF, NARP, LHON): in questi casi è possibile procedere direttamente con la ricerca delle mutazioni puntiformi comunemente associate a tali sindromi, in caso contrario il percorso diagnostico procede con l'esecuzione di indagini specifiche.

Le indagini specifiche comprendono l'analisi istochimica del campione biotico di tessuto muscolare del paziente, l'indagine biochimica e geneticomolecolare. Nell'**analisi istochimica** delle sezioni di tessuto muscolare spesso si evidenzia un caratteristico cambiamento della morfologia muscolare, le cosiddette **ragged red fibers (RRF)**, fibre dall'aspetto stracciato che appaiono con contorni rossi alla colorazione tricromica di Gomori e di colore blu intenso alla colorazione per la succinico deidrogenasi (SDH), per l'accumulo subsarcolemmale di mitocondri; e fibre negative alla colorazione per la citocromo c ossidasi (COX), o **fibre COX**, per riduzione dell'attività del complesso IV della catena respiratoria. La colorazione per l'attività della COX è utilizzata come marker per la funzionalità della catena respiratoria, mentre quella per la SDH evidenzia gli accumuli mitocondriali in quanto unico complesso mitocondriale a intera codifica nucleare.

L'**indagine biochimica** prevede la misura dell'attività dei singoli complessi della catena respiratoria condotta su frazioni od omogenati mitocondriali. I deficit enzimatici possono coinvolgere i singoli complessi o più complessi respiratori. L'esecuzione dei **test geneticomolecolari** è prevista in tutti i casi con anomalie biochimiche e istochimiche, ma anche quando vi sia forte sospetto clinico pur in assenza di anomalie nel campione biotico di muscolo scheletrico. Allo stato attuale le terapie sono meramente sintomatiche, sebbene molto recentemente siano emersi nuovi sviluppi per contrastare i livelli di acido lattico e l'aumento di stress.

Una delle caratteristiche più rilevanti del genoma mitocondriale è l'**eredità uniparentale**, che si osserva nella maggior parte degli eucarioti, inclusi funghi, alghe, piante e vertebrati, e si realizza attraverso meccanismi differenti. Per esempio, nei vertebrati il genoma mitocondriale è ereditato esclusivamente dalla madre. Difatti, l'oocita materno contiene un numero elevato di mitocondri, mentre lo spermatozoo ne contiene pochissimi che vengono anche successivamente degradati nella cellula fecondata.

Anche se in termini di dimensione il genoma mitocondriale (mtDNA) dei vertebrati è circa 200 000 volte più piccolo del genoma nucleare (nDNA), considerando che in una cellula ci possono essere mediamente circa 1000 mitocondri, e quindi circa 10 000 copie di genoma mitocondriale, la frazione di mtDNA ri-

petto a quella di nDNA è significativa e si aggira intorno al 2-3%. L'abbondanza relativa e le ridotte dimensioni rendono particolarmente semplice determinare la sequenza completa del genoma mitocondriale nei metazoi.

Il genoma mitocondriale dei vertebrati ha un tasso di mutazione almeno di un ordine di grandezza più elevato di quello nucleare a causa della prossimità alla catena di trasporto degli elettroni (sede di produzione delle specie reattive dell'ossigeno) e della scarsa efficienza di sistemi di riparazione.

L'elevata frequenza delle mutazioni e la trasmissione uniparentale fanno del genoma mitocondriale un sistema modello molto usato per gli studi di evoluzione e filogenesi molecolare, che ha applicazioni anche nel campo della medicina forense (*vedi* par. 21.8). Gli studi di evoluzione molecolare basati sull'analisi del genoma mitocondriale possono essere condotti sia analizzando i genomi di specie diverse, sia analizzando, in studi di popolazioni, più individui di una stessa specie. Per esempio, l'analisi evolutiva del mtDNA ha contribuito a far luce sulla storia evolutiva sia dei mammiferi (analisi interspecifica) che delle popolazioni umane (analisi intraspecifica) rivelando che l'origine dell'uomo moderno (*Homo sapiens*) è recente (circa 150 000 anni fa) ed è avvenuta in Africa.

5.5 Genoma dei cloroplasti

Nelle piante e negli altri eucarioti fotosintetici, oltre al mitocondrio, è presente un secondo organello, il **cloroplasto** o **plastide**, che avrebbe avuto origine da un cianobatterio endosimbionte (**fig. 5.1**). In aggiunta alla loro funzione centrale nella fotosintesi, i cloroplasti partecipano anche alla biosintesi di amido, acidi grassi, pigmenti e amminoacidi.

Come i mitocondri, anche i cloroplasti posseggono un sistema genetico indipendente; l'apparato per la sua espressione, come la maggior parte dei geni dell'organello endosimbionte originario, è stato progressivamente trasferito al nucleo dell'ospite. La ragione di questo trasferimento massivo è tuttora dibattuta ma si ritiene che esso abbia comportato almeno due significativi vantaggi: (1) i geni plastidici sono esposti a elevati livelli di specie reattive e potenzialmente mutagene generate nel corso dei processi di trasferimento degli elettroni legati alla fotosintesi; (2) la riproduzione sessuata dei geni nucleari offre una maggiore capacità di adattamento e sistemi di riparazione più efficienti della riproduzione asessuata dei genomi degli organelli. Ovviamente le stesse considerazioni valgono per spiegare il trasferimento massivo subito anche dall'endosimbionte da cui ha avuto origine il mitocondrio. Resta tuttavia oscuro il motivo per cui gli organelli conservano un proprio genoma, con tutto l'armamentario necessario alla sua espressione. La spiegazione più semplice che è possibile formulare ipotizza che le funzioni essenziali di questi organelli richiedano che una parte delle proteine debba essere necessariamente generata all'interno dell'organello in quanto la loro elevata idrofobicità non ne renderebbe possibile l'import.

Il genoma dei cloroplasti è una molecola circolare di DNA, di dimensioni comprese tra 70 e 200 kpb, e con una struttura quadripartita che comprende due ripetizioni invertite (IR, *Inverted Repeats*) di dimensioni comprese tra 5 e 76 kpb che contengono i geni per gli rRNA, separate da due regioni a singola copia denominate *Large Single Copy Region* e *Small Single Copy Region* (LSC, SSC). Le differenze di lunghezza sono essenzialmente dovute alle diverse dimensioni delle regioni IR.

Il genoma dei cloroplasti ha normalmente un elevato contenuto in A+T ($\geq 60\%$) e codifica per 150-200 geni coinvolti nell'espressione genica (trascrizione, traduzione), nel metabolismo e nella fotosintesi. I geni plastidici, come quelli dei mitocondri delle piante, possono contenere introni autocatalitici di classe I e II (*vedi* par. 13.1) e i loro trascritti possono essere soggetti a RNA editing, prevalentemente di tipo C $\rightarrow$ U (*vedi* par. 13.2). Il tasso di mutazione dei genomi plastidici è maggiore rispetto a quello del corrispondente genoma nucleare, e anche

di quello mitocondriale. Difatti la velocità di evoluzione dei genomi vegetali ha il seguente trend: cpDNA → nDNA → mtDNA.

Come abbiamo già visto per il genoma mitocondriale, anche il genoma del cloroplasto è presente in più copie in ciascun cloroplasto e vi sono più cloroplasti per cellula. Anche questo è ereditato per via materna, nonostante vi siano casi di ereditarietà paterna o biparentale. Anche il genoma di cloroplasto è molto usato come sistema modello per gli studi di evoluzione e filogenesi molecolare.

5.6 Genoma dei virus

I virus sono parassiti intracellulari obbligati presenti in tutte le forme di vita cellulari, dai batteri all'uomo. Posseggono un proprio genoma, in grado di replicarsi esclusivamente all'interno dell'ospite che infettano, impacchettato all'interno di una struttura proteica detta **capside**. Il genoma virale contiene i geni necessari al suo ciclo vitale che vengono espressi utilizzando gli apparati dell'ospite. Per tale ragione la biologia del genoma virale deve essere compatibile con quella del genoma dell'ospite.

I genomi virali mostrano un'elevatissima variabilità strutturale. Possono essere costituiti da DNA o RNA a singolo o doppio filamento, in forma circolare, lineare o segmentata e dimensioni variabili da poche kpb (fagi batterici) fino a oltre 1 Mpb (mimivirus). I virus a RNA a singolo filamento possono a loro volta essere a filamento (*strand*) positivo (+), quando la sequenza del genoma a RNA ha la stessa polarità dell'RNA che viene tradotto, o negativo (–) nel caso di polarità opposta.

I virus a RNA(+) hanno dimensioni variabili da circa 3,5 kpb (fagi MS2 e QB) a 30 kpb (Coronavirus), possono essere direttamente infettivi senza il concorso di proteine virali e posseggono sequenze non tradotte (UTR, *UnTranslated Region*) al 5' e al 3'. Nei virus eucariotici la sequenza 5'UTR è generalmente modificata dall'aggiunta di un cap (*vedi* par. 13.2) o da una proteina legata covalentemente, mentre la sequenza 3'UTR è normalmente poliadenilata. I virus a RNA(–) devono essere prima ricopiati da una polimerasi per formare una molecola di RNA(+) che potrà essere direttamente tradotta o utilizzata per formare molecole di DNA a doppio filamento, denominate intermedi o forme replicative, che potranno essere usate per successivi cicli di sintesi di mRNA virali.

Il genoma dei virus a DNA è costituito da una o più molecole di DNA, lineari o circolari, di dimensioni variabili dai circa 5 kpb del poliomavirus a oltre 1 Mpb di *Acanthamoeba polyphaga mimivirus*, che codifica per circa un migliaio di proteine e che quindi ha paradossalmente un genoma più grande di alcuni procarioti parassiti.

Una classe particolare di virus è costituita dai viroidi, patogeni delle piante, che posseggono un piccolo genoma circolare a RNA in grado di replicarsi autonomamente nelle cellule vegetali e che non codificano proteine né formano particelle virali o virioni.

Lecture di approfondimento e siti web

La classificazione e la nomenclatura dei batteri attualmente in uso sono accessibili presso il sito LBSN: www.bacterio.cict.fr/

L'elenco completo dei genomi batterici ed eucariotici completamente sequenziati, o in corso di completamento, è accessibile presso il sito GOLD: www.genomesonline.org. Binnewies, T.T., Motro, Y., Hallin, P.F., Lund, O., Dunn, D., La, T., Hampson, D.J., Bellgard, M., Wassenaar, T.M. e Ussery, D.W. (2006), «Ten years of bacterial genome sequencing: comparative-genomics-based discoveries», *Funct. Integr. Genomics.*, **6**(3), pp. 165-185.

Frost, L.S., Leplae, R., Summers, A.O. e Toussaint, A. (2005), «Mobile genetic elements: the agents of open source evolution», *Nat. Rev. Microbiol.*, **3**, pp. 722-732. International Human Genome Sequencing Consortium (2004), «Finishing the eukaryotic sequence of the human genome», *Nature*, **431**(7011), pp. 931-945.



REPLICAZIONE E MANTENIMENTO DEL GENOMA

Tutti gli organismi viventi, sia procarioti che eucarioti, devono replicare il proprio patrimonio genetico per poterlo trasmettere alle cellule figlie. I meccanismi molecolari utilizzati per la replicazione del DNA possono essere diversi: tale varietà è principalmente influenzata dalla struttura del DNA, che, a seconda degli organismi, può essere lineare o circolare, a doppio o a singolo filamento. In ogni caso, la replicazione del DNA è un processo estremamente fedele, e ogni errore nella replicazione di una sequenza nucleotidica del DNA può generare delle mutazioni.

Per replicare il proprio DNA ogni cellula utilizza un complesso di proteine, genericamente chiamato replisoma, i cui componenti sono descritti nel testo in un certo dettaglio strutturale e funzionale. Inoltre, la replicazione del DNA in cellule somatiche inizia in corrispondenza di specifiche sequenze di DNA, definite origini di replicazione. L'accensione delle origini negli eucarioti è accuratamente controllata durante il ciclo cellulare, dato che ogni cromosoma deve essere replicato una volta, e una volta soltanto, a ogni ciclo. Difetti nella replicazione del DNA e nei suoi controlli possono, infatti, alterare la stabilità del genoma e generare anomalie cromosomiche (cap. 6).

Il controllo della stabilità e dell'integrità del genoma è di fondamentale importanza per tutte le cellule. Il DNA (e quindi il genoma di tutti gli organismi) è, infatti, il bersaglio di prodotti del normale metabolismo cellulare e di una serie di agenti chimico-fisici presenti nell'ambiente che inducono sul DNA una varietà di lesioni. Questi danni al DNA devono essere rimossi attraverso una serie di meccanismi di riparazione specifici che dipendono, sostanzialmente, dal tipo di lesione prodotta sul DNA e dall'attivazione di una serie di risposte cellulari, note come meccanismi di sorveglianza o "checkpoint", che hanno il compito di controbilanciare il potenziale mutagenico e citotossico dei danni al DNA (cap. 7).

I genomi, inoltre, devono potersi continuamente evolvere acquisendo nuove sequenze di DNA o riarrangiando quelle esistenti attraverso meccanismi di ricombinazione genetica. Tale processo è mediato da rotture e riunioni di molecole di DNA ed è, quindi, potenzialmente pericoloso per la cellula, che deve sorvegliare finemente tale meccanismo per mantenere sotto controllo la stabilità del proprio genoma (cap. 8).

6

Replicazione del DNA

6.1 Replicazione semiconservativa del DNA

6.2 Modello del replicone

6.3 Identificazione delle origini di replicazione

6.4 Meccanismo della replicazione del DNA nei procarioti

6.5 Meccanismo della replicazione del DNA negli eucarioti

6.6 Replicazione dei telomeri nei cromosomi lineari degli eucarioti

6.7 Controllo della replicazione del DNA durante il ciclo cellulare

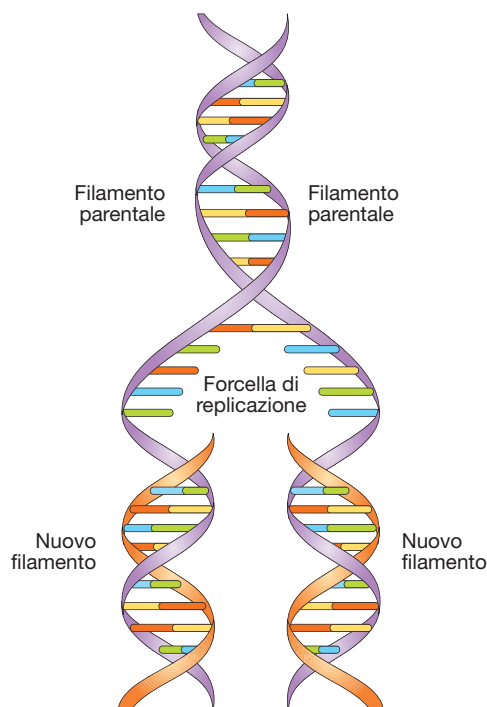
6.8 Alterazioni nei meccanismi di replicazione possono provocare cambiamenti quantitativi di sequenze di DNA nel genoma

James Watson e Francis Crick concludono il loro famoso articolo su *Nature*, in cui nel 1953 propongono la struttura a doppia elica del DNA, con la seguente frase: «It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material». La caratteristica del DNA che più li aveva colpiti era, infatti, la complementarità tra le basi azotate nei due filamenti che costituiscono la doppia elica del DNA, che suggerisce un possibile modello di replicazione (*vedi* cap. 2).

A ogni divisione cellulare, l'informazione genetica contenuta nel DNA deve essere duplicata e copiata con la massima fedeltà possibile per trasmettere alle cellule figlie lo stesso patrimonio genetico della cellula madre. Ogni variazione nella sequenza di nucleotidi nel DNA genera dei cambiamenti, definiti **mutazioni**. Tali cambiamenti sono necessari, in quanto il loro accumularsi nell'arco di milioni di anni è stato alla base del processo evolutivo che ha portato alla variabilità genetica tra specie animali e vegetali e alla diversità degli individui all'interno della stessa specie. Le mutazioni possono però essere dannose e, pertanto, un tasso di mutazione troppo elevato può essere molto pericoloso. È quindi estremamente importante che la replicazione del materiale genetico, cioè del DNA, avvenga nel modo più accurato possibile limitando l'insorgenza di mutazioni durante il processo replicativo (**fig. 6.1**).

Molecole organiche semplici, come gli amminoacidi o i nucleotidi, possono associarsi per formare dei polimeri, e un filamento di DNA altro non è che un polimero di nucleotidi, o **polinucleotide**. In condizioni prebiotiche tali polimeri possono essersi formati casualmente, per esempio quando i composti organici venivano in contatto in un ambiente condizionato dalle alte temperature. Una volta che un polimero si è formato, può influenzare la formazione di altri polimeri, e questo è particolarmente vero per i polinucleotidi, che possono agire da **stampo** in reazioni di polimerizzazione. Un polinucleotide, per esempio un poli-T, può servire da stampo per la sintesi di un secondo polimero composto da un altro nucleotide, per esempio poli-A. Questa capacità dei polimeri di agire

Figura 6.1 La replicazione del DNA avviene in modo semiconservativo. I due filamenti della molecola di DNA parentale (in viola) si separano e servono da stampo per la neosintesi di due filamenti complementari (in arancione). La forcella di replicazione così generatasi può propagarsi lungo il DNA non ancora replicato dando origine a due molecole figlie uguali alla molecola originaria e costituite da un filamento parentale e da uno di neosintesi.



da superfici stampo è ben nota ai chimici e biochimici e dipende dal fatto che, nell'esempio precedente, un poli-T forma legami chimici preferenziali con un poli-A. La capacità dei polinucleotidi di appaiarsi tra di loro ha sicuramente svolto un ruolo cruciale nell'origine della vita.

6.1 Replicazione semiconservativa del DNA

Ritornando alla struttura del DNA determinata da Watson e Crick, la complementarietà delle basi azotate nei due filamenti della doppia elica suggerì loro che ciascun filamento potesse agire come stampo per la sintesi di un filamento di neosintesi complementare al filamento parentale. Questa ipotesi di replicazione del DNA venne definita **replicazione semiconservativa** e prevedeva che i due filamenti parentali si separassero durante il processo replicativo.

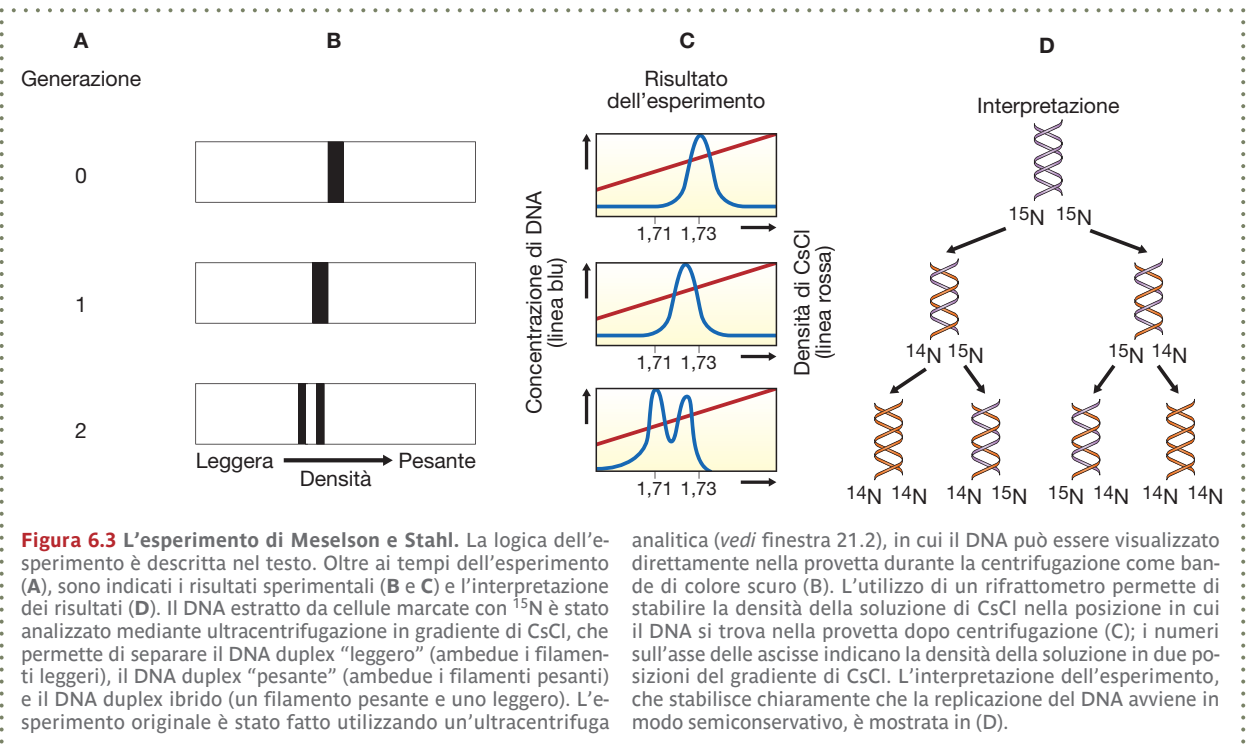
Un altro potenziale meccanismo prevedeva, invece, che i due filamenti della doppia elica parentale rimanessero appaiati e, in un qualche modo, l'intera doppia elica potesse servire da stampo durante il processo replicativo dando origine a una molecola figlia di neosintesi uguale alla molecola originale. Questa ipotesi prendeva il nome di **replicazione conservativa**. Un terzo possibile meccanismo (che sin dall'inizio apparve però molto poco probabile) ipotizzava una replicazione del DNA di tipo dispersivo: durante la replicazione, il DNA verrebbe frammentato e nelle due molecole figlie finirebbero per coesistere tratti di DNA parentali insieme a segmenti di DNA neosintetizzato.



Figura 6.2 Matthew Meselson e Franklin Stahl. Sono gli autori del fondamentale esperimento, pubblicato nel 1958, con cui hanno dimostrato che il DNA si replica in maniera semiconservativa. (Fonte: F.L. Holmes, 1996.)

L'esperimento di Meselson e Stahl

La dimostrazione sperimentale che la replicazione del DNA avviene davvero in modo semiconservativo è stata ottenuta ben 5 anni dopo la scoperta della struttura del DNA. Nel 1958, infatti, Matthew Meselson e Franklin Stahl (**fig. 6.2**) pubblicarono quello che è stato definito come "uno degli esperimenti più eleganti nella storia della Biologia Molecolare" (**fig. 6.3**). Come sempre accade in ambito scientifico, tale risultato è stato possibile perché, nel frattempo, erano state sviluppate nuove tec-



nologie per studiare e analizzare il DNA. Si scoprì, infatti, che il DNA poteva essere marcato e seguito sperimentalmente introducendo nella doppia elica degli isotopi pesanti dell'azoto, un atomo presente nelle basi puriniche e pirimidiniche del DNA. Ciò si può facilmente ottenere facendo crescere dei batteri in un terreno di coltura a cui vengono aggiunti dei sali d'ammonio (quale il cloruro d'ammonio, NH_4Cl) che contengano l'isotopo pesante dell'azoto ^{15}N , al posto dell'isotopo normale ^{14}N .

Molecole di DNA che contengono ^{15}N hanno una "densità" maggiore rispetto a quella di molecole di DNA contenenti ^{14}N e, dopo centrifugazione all'equilibrio in soluzioni concentrate di cloruro di cesio (vedi cap. 20), si troveranno verso il fondo della provetta da centrifugazione, proprio perché hanno densità più elevata.

Dove si trova il DNA nella provetta è facilmente determinabile: è sufficiente bucare la provetta con un ago e raccogliere goccia a goccia il contenuto della provetta in diverse frazioni separate. Il DNA assorbe la luce nell'ultravioletto a 260 nm e quindi, analizzando ciascuna frazione con uno spettrofotometro, si può determinare in quali frazioni si trova il DNA e con un semplice strumento si può anche misurare la densità delle frazioni stesse. Se la quantità di DNA è sufficientemente abbondante, proprio perché il DNA assorbe la luce UV a 260 nm, si può direttamente identificare dove si trova il DNA nella provetta illuminando la provetta stessa con una lampada che emette luce UV (fig. 6.3B).

Meselson e Stahl hanno affrontato il problema della replicazione del DNA utilizzando il batterio *Escherichia coli* come modello sperimentale. Infatti, *E. coli* può facilmente crescere in laboratorio in condizioni di coltura sperimentalmente controllate. Cellule di *E. coli* sono state fatte duplicare per molte generazioni in un terreno contenente sali pesanti dell'azoto, affinché tutte le cellule avessero molecole di DNA contenenti ^{15}N . A questo punto, le cellule sono state raccolte tramite centrifugazione, lavate per togliere ogni traccia di terreno contenente ^{15}N , diluite in un terreno contenente sali d'ammonio con l'isotopo leggero ^{14}N e fatte crescere in queste condizioni per due generazioni.

L'ipotesi sperimentale di Meselson e Stahl era che, se fosse stata vera l'ipotesi semiconservativa, dopo una generazione tutte le cellule di *E. coli* che avevano

replicato il loro DNA avrebbero dovuto avere molecole di DNA contenente un filamento parentale marcato con ^{15}N e un filamento neosintetizzato marcato con ^{14}N . Tutte le molecole di DNA avrebbero dovuto avere, quindi, una densità intermedia rispetto a quella di un DNA a doppia elica tutto pesante o tutto leggero. Alla seconda generazione, metà del DNA avrebbe dovuto avere una densità intermedia e metà una densità corrispondente al DNA leggero. In ogni caso, sin dalla prima generazione, non si sarebbe mai dovuto trovare DNA completamente pesante.

Al contrario, se fosse risultata vera l'ipotesi conservativa della replicazione del DNA, i ricercatori si aspettavano che alla prima generazione metà del DNA sarebbe stato pesante e metà del DNA sarebbe stato leggero, mentre non si sarebbe mai dovuto trovare DNA con una densità intermedia.

I risultati sperimentali ottenuti da Meselson e Stahl furono inequivocabili e dimostrarono per la prima volta che la replicazione del DNA di *E. coli* avviene in modo semiconservativo. Poiché, come ebbe a dire alcuni anni dopo Jacques Monod, "... quello che è vero per *E. coli* è vero anche per un elefante ...": la replicazione del DNA avviene in modo semiconservativo in tutti gli organismi, siano essi procarioti o eucarioti.

6.2 Modello del replicone

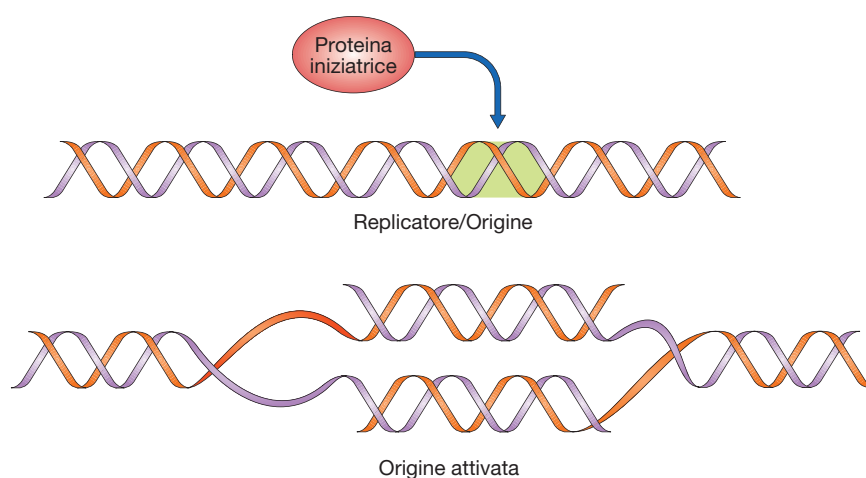
Come si realizza a livello molecolare la replicazione semiconservativa del DNA? Dopo oltre 50 anni dalla scoperta della struttura del DNA conosciamo molti aspetti del suo meccanismo di replicazione, ma altri restano ancora da chiarire a dimostrazione della complessità del processo.

Gli aspetti che più hanno catturato la curiosità dei ricercatori sin dalla scoperta del DNA e la definizione della sua struttura sono stati essenzialmente tre:

1. L'inizio della replicazione del DNA avviene a caso o parte da specifiche sequenze nucleotidiche?
2. Quali sono le proteine che realizzano il processo replicativo e come si assembla l'intera macchina proteica coinvolta in questo processo?
3. Come si integra la replicazione del DNA con il meccanismo di divisione cellulare, dato che ogni cellula figlia deve ereditare il patrimonio genetico contenuto nella cellula madre?

Tutti e tre i quesiti furono affrontati da François Jacob, Sydney Brenner e François Cuzin in un famoso articolo del 1963 (10 anni dopo la scoperta della struttura del DNA!) in cui proposero il modello del **replicone**. Tale modello postula che l'inizio della replicazione del DNA è geneticamente controllato da

Figura 6.4 Modello del "replicone". Il modello suggerisce che sequenze sul DNA (che agiscono quindi in *cis*), originariamente chiamate "replicatori", determinano dove debba iniziare la replicazione del DNA. Oggi queste sequenze di DNA vengono chiamate "origini di replicazione". Con l'origine di replicazione interagiscono proteine (che agiscono quindi in *trans*) genericamente qui definite "proteina iniziatrice". Il "replicone" è l'unità di replicazione e ciascuna molecola di DNA può contenere uno o più repliconi.



sequenze specifiche in *cis* sul DNA (chiamate **replicatori**). Queste sequenze determinano dove può partire la replicazione del DNA interagendo con specifiche proteine (**iniziatori**) che agiscono in *trans* e collegano il processo replicativo con la crescita e la divisione cellulare (**fig. 6.4**).

Il modello del replicone fu ispirato dai dati ottenuti da John Cairns, che analizzò la geometria del cromosoma di *E. coli* in fase replicativa utilizzando una tecnica nota come autoradiografia. *E. coli* contiene un'unica molecola di DNA circolare di grandi dimensioni, circa 4,2 Mpb o circa 1300 μm . È possibile rendere radioattivo il DNA di *E. coli* facendo crescere le cellule batteriche in un terreno di coltura che contiene $[^3\text{H}]$ -timidina che viene incorporata nel DNA; le cellule batteriche vengono poi lisate delicatamente su un vetrino da microscopio su cui viene stesa un'emulsione fotosensibile. La radioattività emessa dal trizio produce un'immagine del cromosoma visibile al microscopio ottico (**fig. 6.5A**).

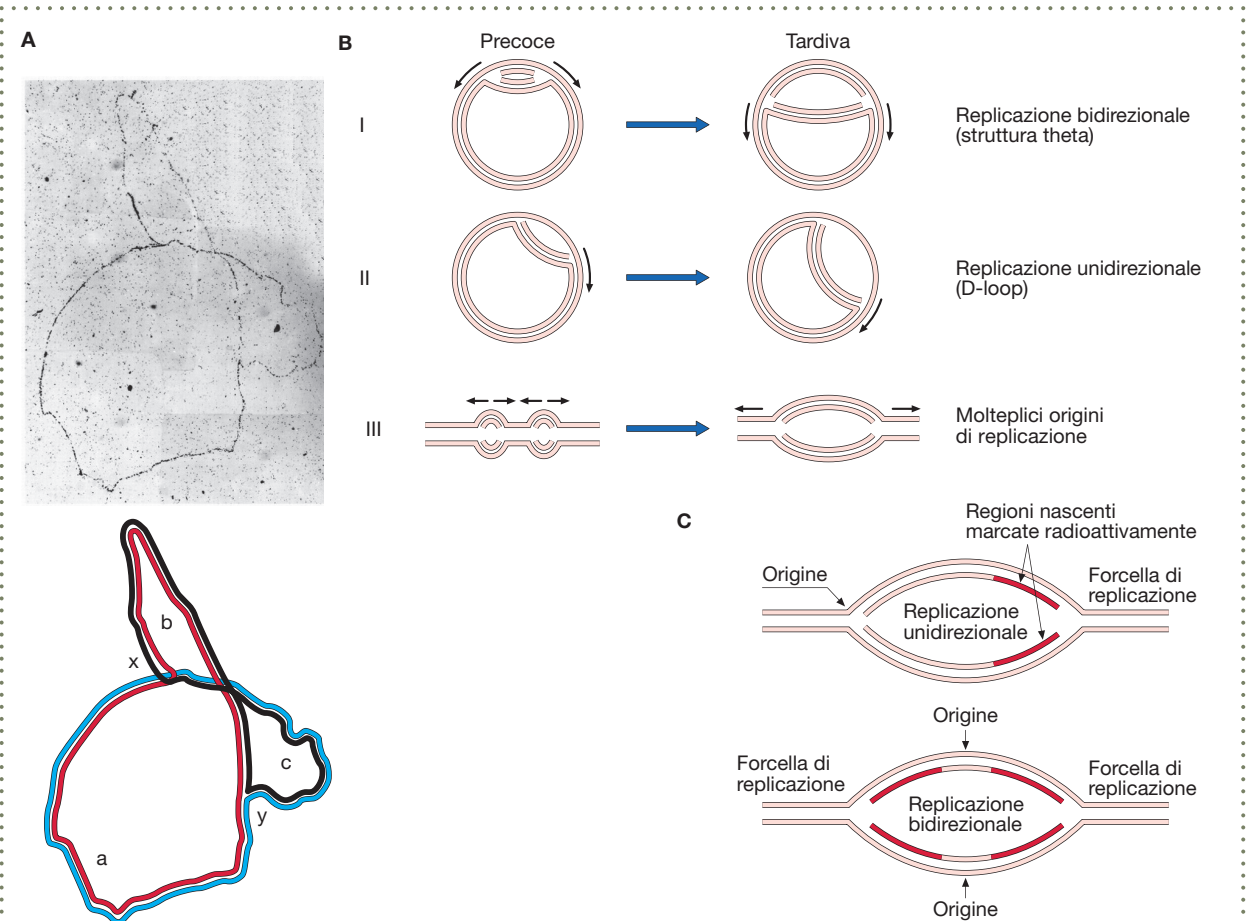


Figura 6.5 Replicazione del DNA del batterio *Escherichia coli* e modelli di replicazione. (A) L'immagine autoradiografica, tratta dal lavoro originale di J. Cairns, mostra la molecola del DNA di *E. coli* in fase di replicazione. In basso è disegnato un diagramma interpretativo del risultato dell'autoradiografia. La logica dell'esperimento e il risultato sono spiegati nel testo. Le due forcelle di replicazione (x e y) creano sul DNA circolare di *E. coli* le tre anse a, b e c. L'ansa a corrisponde alla parte della molecola che è stata replicata due volte in terreno radioattivo; le anse b e c, che sono di uguale lunghezza, corrispondono alle regioni che hanno replicato una sola volta e hanno quindi un solo filamento radioattivo. Infatti l'ansa a ha una densità di granuli doppia rispetto alle anse b e c. La linea nera indica il DNA iniziale non radioattivo, le linee azzurra e rossa i filamenti marcati radioattivamente. (B) Sono indicate le strutture che si formano, a stadi precoci e tardivi, durante la

replicazione di una molecola di DNA a doppio filamento circolare in cui esiste una sola origine di replicazione e questa procede bidirezionalmente (I) o unidirezionalmente (II). Nello schema III è mostrata la struttura di una molecola di DNA a doppio filamento lineare contenente due origini di replicazione che, procedendo bidirezionalmente, si fondono una con l'altra a tempi tardivi. (C) Se si marca con timidina radioattiva il DNA di nuova sintesi e si analizzano le regioni nascenti tramite autoradiografia (vedi fig. 6.7), sono prevedibili risultati diversi se la sintesi del DNA procede unidirezionalmente (in alto) o bidirezionalmente (in basso). I filamenti di DNA di neosintesi sono indicati in rosso. Si noti che la marcatura appare non iniziare all'origine ma un po' dopo; ciò dipende dal fatto che le cellule, non essendo sincronizzate, possono avere già cominciato la replicazione del DNA al momento dell'inizio della marcatura.

La tecnica è molto complessa e ottenere delle molecole integre di DNA di *E. coli* è alquanto difficile. Le immagini ottenute possono però chiarire alcuni parametri importanti del processo replicativo. Il cromosoma di *E. coli* in fase di replicazione mostrato in **figura 6.5A** permette di visualizzare tre parti della molecola: **a**, **b** e **c**. Si può notare che la grande ansa **a** presenta una densità di granuli radioattivi sulla lastra fotografica circa doppia di quella osservabile nelle anse **b** e **c**. Nell'esperimento mostrato, cellule di *E. coli* cresciute in terreno non radioattivo sono state fatte replicare per un'intera generazione e parte di una seconda in presenza di timidina radioattiva. La porzione di DNA che ha replicato una sola volta in presenza di [^3H]-timidina (**b** e **c**) possiede un filamento marcato (linee rossa e azzurra) e un filamento non marcato (linea nera). Il DNA che ha replicato due volte in presenza di [^3H]-timidina (**a**) possiede entrambi i filamenti marcati (*vedi* didascalia della **figura 6.5A**). Come atteso, le porzioni del cromosoma in cui entrambi i filamenti sono stati sintetizzati in presenza di timidina radioattiva (**a**) hanno una densità di grani doppia rispetto alle porzioni in cui uno solo dei filamenti è stato sintetizzato in presenza di [^3H]-timidina (**b** e **c**). Questo risultato, tra l'altro, conferma che la replicazione del cromosoma di *E. coli* avviene in modo semiconservativo. L'intermedio di replicazione mostrato in **figura 6.5A** assomiglia, in una proiezione planare (**fig. 6.5B**), alla lettera greca θ ed è stato perciò chiamato "struttura theta".

Il punto d'inizio della replicazione viene oggi definito **origine di replicazione**. Come mostrato in **figura 6.5B**, per una molecola di DNA circolare, la replicazione può procedere bidirezionalmente originando delle "bolle" di replicazione sempre più grandi (I). La replicazione bidirezionale a partire da specifiche origini è il meccanismo più comunemente usato in procarioti ed eucarioti, anche se in alcuni casi la replicazione può procedere unidirezionalmente (II). Mediante autoradiografia è anche possibile valutare se su un'unica molecola di DNA esistono più origini di replicazione (III): numerose origini su uno stesso cromosoma sono davvero utilizzate per replicare le lunghe molecole di DNA lineari presenti negli eucarioti. L'autoradiografia permette anche di distinguere se una bolla di replicazione si estende in una o in entrambe le direzioni (**fig. 6.5C**).

6.3 Identificazione delle origini di replicazione

Come accennato precedentemente, un'origine di replicazione corrisponde a una particolare sequenza di DNA da cui parte il processo replicativo, che procede nella maggior parte dei casi bidirezionalmente originando due forcelle di replicazione (**fig. 6.5**). Molecole di DNA relativamente semplici, come quelle dei plasmidi e della maggior parte dei batteriofagi e dei virus, posseggono una singola origine di replicazione per molecola di DNA. La presenza di un'origine di replicazione in questi elementi genetici è essenziale perché DNA di plasmidi, batteriofagi o virus siano in grado di replicarsi autonomamente all'interno delle cellule procariotiche o eucariotiche in cui si moltiplicano.

Origine di replicazione di *Escherichia coli*

Il cromosoma di *E. coli*, pur avendo una dimensione di 4,2 Mpb, viene replicato utilizzando una singola origine. L'origine di replicazione di *E. coli* è chiamata **oriC** ed è stata identificata e localizzata utilizzando strategie diverse. Una di queste strategie è stata escogitata poco dopo che si sono sviluppate le tecniche per costruire molecole di DNA ricombinanti (*vedi* cap. 20). Come mostrato in **figura 6.6**, si è partiti da un tipico plasmide di *E. coli* contenente un gene che conferisce resistenza all'antibiotico ampicillina e, mediante il taglio con uno specifico enzima di restrizione, è stata eliminata la porzione del plasmide contenente la sua origine di replicazione. Questo plasmide difettivo, incapace di replicarsi in *E. coli*, è stato poi utilizzato per costruire una banca di DNA di *E. coli*: l'ipotesi di lavoro era che

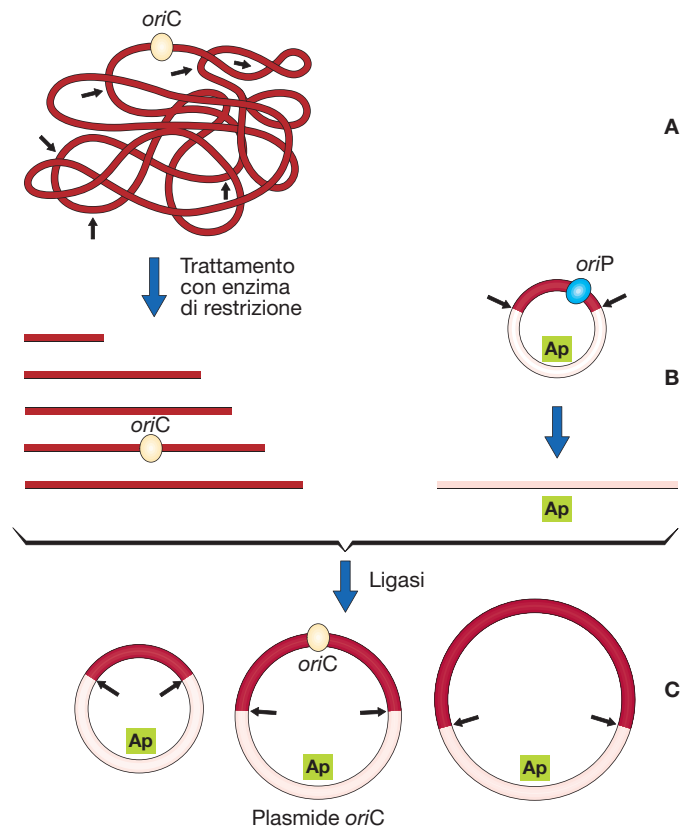


Figura 6.6 Schema della procedura utilizzata per individuare la sequenza di DNA corrispondente all'origine di replicazione nel cromosoma circolare di *Escherichia coli* (*oriC*). (A) Il cromosoma circolare di *E. coli* è schematicizzato in rosso con indicata un'ipotetica posizione per la sua origine di replicazione (*oriC*). (B) Il DNA di *E. coli* è stato tagliato in numerosi frammenti con un enzima di restrizione; i tagli sono indicati in (A) dalle frecce nere. Solo uno dei frammenti di restrizione generati contiene *oriC*. Lo stesso enzima di restrizione provoca due tagli in un plasmide contenente un marcatore genetico in grado di conferire resistenza all'antibiotico ampicillina (Ap) e la sua origine di replicazione (*oriP*), permettendo

così la rimozione dell'origine dal plasmide stesso. I frammenti del cromosoma di *E. coli* vengono clonati nel plasmide privato della sua origine ma contenente il segmento di DNA che conferisce la resistenza all'ampicillina. (C) Delle varie molecole di DNA ricombinante ottenute, solo quelle in cui si è inserita *oriC*, come la molecola al centro, sarà in grado di replicarsi dopo trasformazione in cellule di *E. coli*, in quanto possiede *oriC*, che sostituisce funzionalmente l'origine di replicazione del plasmide rimossa. I batteri trasformati contenenti frammenti di DNA con *oriC* possono venire selezionati positivamente, in quanto in grado di crescere in un terreno di coltura addizionato con ampicillina (vedi cap. 20).

frammenti di DNA di *E. coli* contenenti *oriC* potessero riconferire al plasmide difettivo la capacità di replicarsi e quindi, alla cellula che lo ospita, la possibilità di crescere in un terreno contenente ampicillina (fig. 6.6). Si è visto, poi, che plasmidi contenenti *oriC* richiedono le stesse proteine necessarie per la replicazione del DNA di *E. coli*, suggerendo che potevano essere utilizzati come modelli corretti per lo studio funzionale dell'origine di replicazione di un batterio.

Una volta isolato un frammento di DNA contenente *oriC* si è provveduto ad accorciarlo sempre di più per identificare quale fosse la più corta sequenza di DNA di *E. coli* in grado di funzionare come origine. Ci si è domandati, inoltre, cosa avesse di speciale questa regione di DNA rispetto al resto del cromosoma di *E. coli*. Questa è la tipica domanda che i ricercatori si pongono quando vogliono capire la funzione di una specifica sequenza di DNA in processi quali la replicazione, la trascrizione, la ricombinazione ecc. Nel caso di *oriC* è stato osservato che l'origine "minima" di *E. coli* è piuttosto piccola, solo 245 pb. È una regione più ricca in A e T della restante parte del cromosoma di *E. coli* e il suo sequenziamento ha permesso di stabilire che contiene 4 ripetizioni di 9 pb e 3 di 13 pb molto simili, rispettivamente, tra di loro (vedi oltre, fig. 6.18A).

La particolarità di essere ricche in A e T è poi risultata, come vedremo, una caratteristica comune alle origini di replicazione identificate in altri organismi ed è legata al fatto che un'origine di replicazione, per funzionare, deve potersi aprire (**fig. 6.5B**) ed è necessaria meno energia per denaturare localmente una regione di DNA ricca in A e T piuttosto che in C e G (*vedi* cap. 20). Quando una regione di DNA svolge una specifica funzione, e in essa si trovano corte sequenze (spesso chiamate **box**) con composizione simile e/o ripetute più volte, è possibile ipotizzare che tali sequenze possano servire come un segnale di riconoscimento. Come vedremo più avanti, le corte sequenze ripetute presenti in *oriC* vengono infatti riconosciute da proteine coinvolte nelle fasi iniziali della replicazione del DNA di *E. coli*.

Specifiche origini di replicazione sono state individuate in plasmidi, fagi e virus di eucarioti, così come nel DNA presente in alcuni organelli cellulari, quali i mitocondri. Le caratteristiche generali sono sempre di essere ricche in A e T e contenere sequenze riconosciute da proteine che agiscono come iniziatori della replicazione. La differenza principale rispetto alla replicazione del cromosoma di *E. coli* è che alcune di queste molecole di DNA vengono replicate unidirezionalmente o, almeno, in modo asimmetrico, per cui le due forcelle di replicazione (*replication forks*) procedono in modo asincrono.

Origini di replicazione negli eucarioti

L'esistenza di origini di replicazione negli eucarioti, soprattutto nei metazoi, ha dato origine a intensi dibattiti. Infatti, durante i primi stadi dello sviluppo embrionale in *Drosophila melanogaster* e in *Xenopus laevis*, non sembrano esistere siti specifici d'inizio della replicazione del DNA, ma la sintesi di DNA parte essenzialmente da sequenze casuali. Per di più, se si inietta del DNA eterologo, per esempio DNA di un batteriofago, in un oocita di *Xenopus*, il DNA eterologo viene replicato, indicando l'assenza di una specie-specificità della replicazione del DNA nelle cellule embrionali. Durante le prime divisioni cellulari dello zigote la necessità principale sembra quella di replicare il DNA il più rapidamente possibile, concludendo velocemente la divisione cellulare. Solo dopo lo stadio di blastula la divisione cellulare rallenta e la replicazione non parte più a caso ma da specifiche origini. Quali siano i fattori che determinano tale transizione non è ancora completamente chiarito: si sa che la trascrizione svolge un ruolo importante, e altre condizioni rilevanti possono essere cambiamenti nelle concentrazioni dei livelli di nucleotidi, la quantità di specifiche proteine iniziatrici, la struttura della cromatina e modificazioni epigenetiche degli istoni.

Quello che è certo è che in tutti i cromosomi degli eucarioti esistono numerose origini di replicazione. Non solo, ma non tutte le sequenze che potenzialmente possono agire come origini di replicazione vengono utilizzate a ogni ciclo di divisione cellulare.

L'esistenza di molteplici origini di replicazione negli eucarioti è stata per la prima volta dimostrata da Joel A. Huberman e Arthur Riggs, mediante esperimenti autoradiografici (schematizzati in **fig. 6.7B** mentre un esperimento originale è mostrato in **fig. 6.7C**) ed è anche osservabile mediante microscopia elettronica (**fig. 6.7D**). L'osservazione di "bolle" di replicazione di dimensioni diverse (**fig. 6.7A e D**) indica immediatamente due caratteristiche importanti delle origini dei cromosomi degli eucarioti: (1) l'accensione delle origini non è simultanea, ma esistono delle origini "*early*" (precoci) che vengono "accese" presto durante la fase di replicazione del DNA (fase S), mentre altre origini "*late*" (tardive) vengono attivate a tempi successivi; (2) le forcelle di replicazione che avanzano in direzioni opposte sul cromosoma possono fondersi tra di loro (**fig. 6.7A**).

Ogni origine di replicazione individua un possibile "replicone". Va però sottolineato che non tutte le origini di replicazione sono necessariamente utilizzate durante la replicazione dei genomi degli eucarioti. Sorprendentemente, quindi,

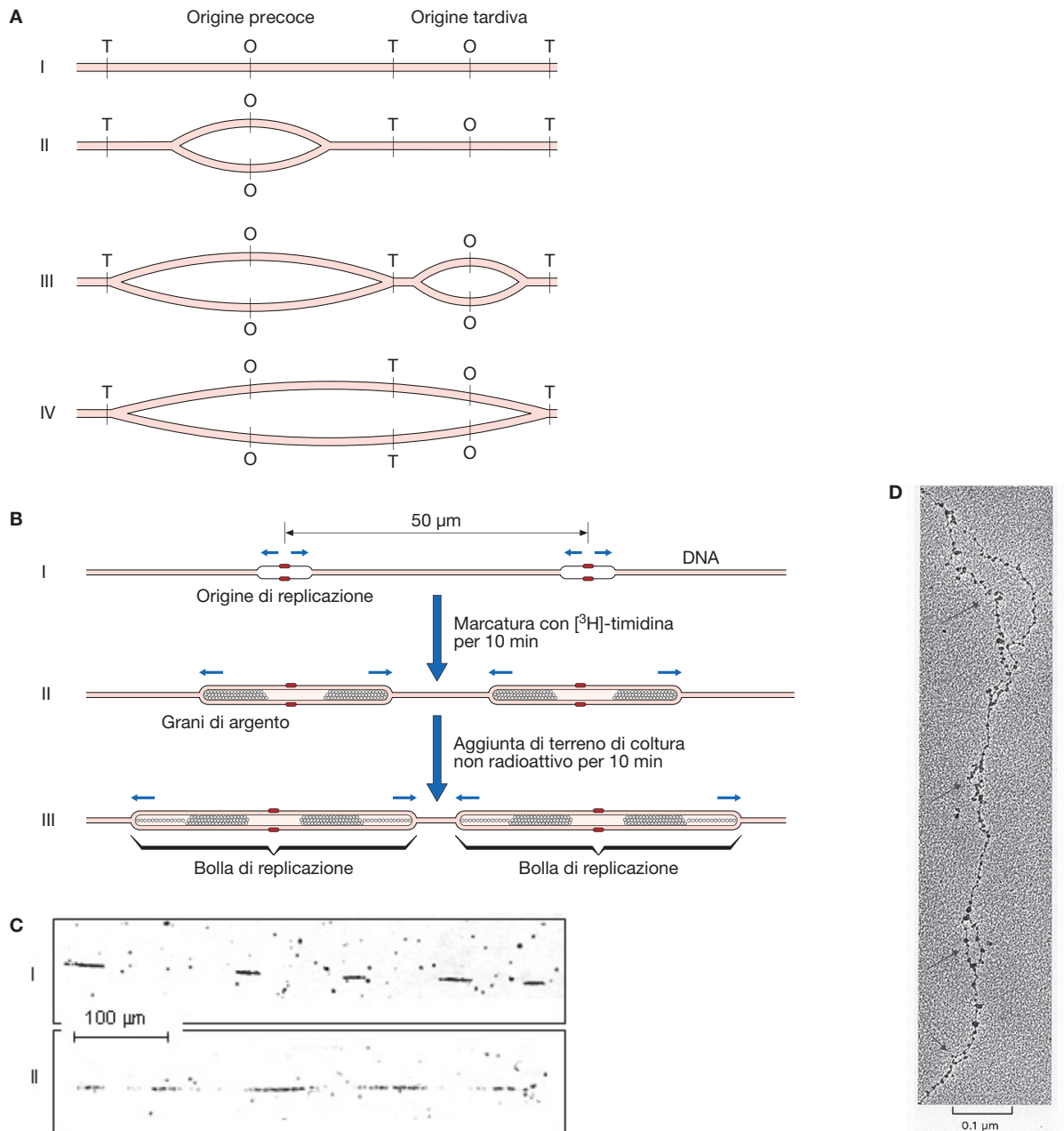


Figura 6.7 Ciascun cromosoma degli organismi eucariotici contiene molteplici origini di replicazione. (A) Il DNA dei cromosomi eucariotici è organizzato in diverse unità di replicazione (repliconi). Ognuna di queste è caratterizzata da un punto di origine (O) da cui partono due forcelle di replicazione che procedono bidirezionalmente. I punti di terminazione (T) di due repliconi adiacenti coincidono, così che, a tempi tardivi di replicazione, repliconi adiacenti si fondono uno nell'altro. L'accensione delle origini di replicazione non è simultanea, ma alcune (precoci o *early*) vengono accese durante la fase S prima di altre (tardive o *late*). (B) L'esistenza di tale organizzazione dei repliconi negli eucarioti è stata originariamente dimostrata mediante autoradiografia da Joel Huberman e Arthur Riggs. La procedura è simile a quella descritta per la figura 6.5. Brevemente, in cellule umane in coltura non sincronizzate il DNA è già parzialmente replicato al momento dell'inizio della marcatura (I); il DNA è stato poi marcato per soli 10 minuti con $[^3\text{H}]$ -timidina altamente radioattiva (II), oppure per ulteriori 10 minuti con $[^3\text{H}]$ -timidina

molto meno radioattiva. L'incorporazione del precursore (III) radioattivo viene visualizzata coprendo il vetrino su cui sono state deposte le molecole di DNA con un'emulsione fotografica contenente bromuro d'argento. Le radiazioni emesse dalla timidina radioattiva incorporata nel DNA colpiscono l'emulsione causando la deposizione di granuli di argento metallico, che sono visualizzabili mediante autoradiografia e analisi al microscopio ottico. Si noti che la distribuzione dei granuli nello schema dimostra non solo la presenza di due origini sulla fibra di DNA in esame, ma anche che le forcelle di replicazione procedono bidirezionalmente. (C) Sono qui mostrati i risultati ottenuti in esperimenti di autoradiografia del DNA condotti come descritto in (B). (D) È possibile "cattare" direttamente, mediante microscopia elettronica, immagini di DNA o di cromatina in replicazione. Nella fotografia è osservabile un tratto di cromatina che mostra quattro "bolle" di replicazione indicate dalle frecce. La dimensione diversa delle bolle suggerisce che le più estese sono state accese prima di quelle più piccole.

il replicone più grande è quello di *E. coli*: infatti, dato che nel genoma di tale organismo esiste una sola origine di replicazione, la dimensione del replicone di *E. coli* coincide con l'intero genoma (4,2 Mpb/1). La replicazione di un replicone così grande fa sorgere il quesito di come una molecola di 4,2 Mpb possa replicarsi nel breve tempo (circa 20 minuti) richiesto per la divisione di una cellula di *E. coli*. La risposta a questa domanda è da ricondursi alla velocità del processo replicativo, che è molto più alta nei procarioti rispetto agli eucarioti. In *E. coli* la forcella di replicazione si estende alla velocità di circa 50 000 pb/minuto, mentre negli eucarioti tale velocità varia, a seconda delle specie, da 500 a 3500 pb/minuto. Dal lievito ai mammiferi il numero delle origini di replicazione per genoma è compreso tra alcune centinaia (lievito) a 10-30 000 (mammiferi) e tale differenza è legata alle diverse dimensioni del genoma. Tuttavia, la lunghezza media dei repliconi è abbastanza costante in tutti gli eucarioti (40-100 kpb). Nel corso dell'evoluzione, il passaggio da organismi con una sola origine di replicazione (procarioti) a organismi con più origini di replicazione (eucarioti) ha rappresentato una transizione importantissima e probabilmente essenziale per l'evoluzione di genomi di grandi dimensioni.

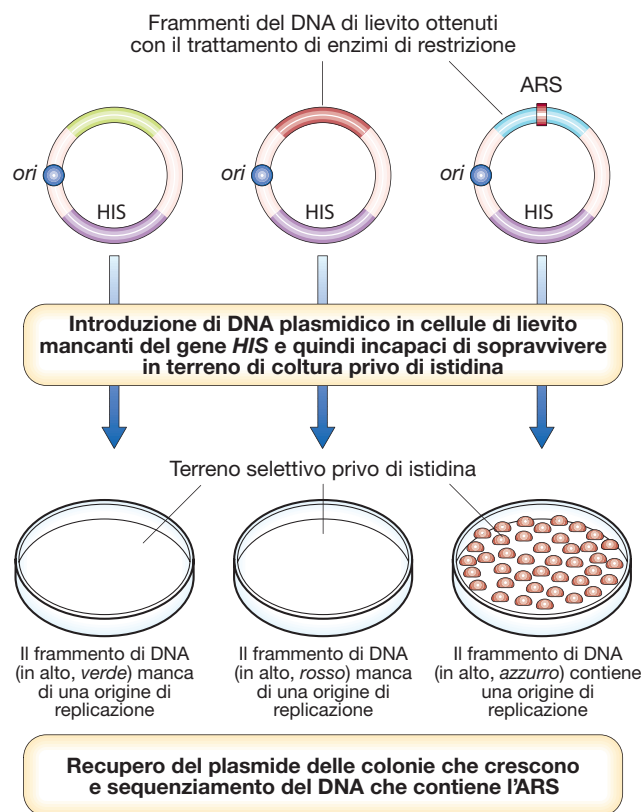


Figura 6.8 Identificazione delle origini di replicazione (*Autonomously Replicating Sequences* o *ARS*) nel lievito *Saccharomyces cerevisiae*. L'identificazione delle origini di replicazione presenti nel genoma di lievito si è basata sul presupposto che un frammento di DNA contenente un'origine di replicazione di lievito deve essere in grado di conferire a un plasmide la capacità di replicarsi in cellule di lievito. Analogamente a quanto descritto in figura 6.6 per il clonaggio di *oriC* di *E. coli*, si è proceduto a creare una libreria di frammenti di DNA di lievito, ottenuti mediante digestione del suo genoma con un enzima di restrizione, all'interno di un vettore plasmidico che è in grado di replicarsi in cellule di *E. coli* perché con-

tiene la sua propria origine (*ori*), ma che non si replica in cellule di lievito. Il vettore contiene anche un marcatore, qui indicato come *HIS*, perché se il plasmide è in grado di replicarsi, una volta entrato all'interno di una cellula di lievito, la rende capace di crescere in un terreno privo dell'amminoacido istidina. Infatti, il ceppo di lievito utilizzato per la trasformazione della libreria ha una mutazione in uno dei geni richiesti per la biosintesi di tale amminoacido. Come indicato nella parte in basso della figura, solo le cellule di lievito che, dopo trasformazione, hanno un plasmide contenente un frammento di DNA con una *ARS* saranno in grado di crescere su terreno selettivo privo di istidina, generando colonie.

ARS del lievito *Saccharomyces cerevisiae*

Le prime origini di replicazione di eucarioti sono state caratterizzate nel lievito *Saccharomyces cerevisiae*, che è uno dei sistemi biologici modello preferiti dai genetisti e dai biologi molecolari (vedi cap. 22). Il lievito è, infatti, un eucariote a tutti gli effetti, ma ad esso possono essere applicate tecniche di selezione e condizioni di coltura analoghe a quelle utilizzate per i batteri. Per quanto riguarda la replicazione del DNA, il lievito ha un genoma (13,5 Mpb) solo 2,5 volte più grande di quello di *E. coli*, suddiviso in 16 cromosomi lineari, ognuno contenente numerose origini di replicazione. Le origini di replicazione di *S. cerevisiae* sono state individuate e localizzate con una logica molto simile a quella descritta precedentemente per l'identificazione di *oriC* (fig. 6.8).

Un plasmide che normalmente si propaga in cellule di *E. coli* non è in grado di replicarsi autonomamente se introdotto in cellule di *S. cerevisiae*. Questo è in qualche modo atteso, dato che il plasmide contiene un'origine di replicazione riconoscibile da proteine presenti in *E. coli*, ma non in *S. cerevisiae*, che utilizza, invece, un corredo di proteine specifiche per il riconoscimento delle sue proprie origini (vedi oltre). È pertanto possibile tentare di costruire una banca di DNA di lievito in un plasmide batterico e valutare se specifici frammenti di DNA di lievito, inseriti nel plasmide mediante classiche metodologie di DNA ricombinante (vedi cap. 20), lo rendano capace di replicarsi una volta introdotto all'interno di una cellula di lievito. Questa tecnica è facilmente applicabile in tale organismo perché si possono utilizzare cellule di lievito mutanti che richiedono la presenza e la replicazione del plasmide per crescere su uno specifico terreno selettivo (fig. 6.8).

Se un frammento di DNA di lievito contiene un'origine di replicazione può essere individuato all'interno della banca di DNA perché tale origine rende il plasmide capace di replicarsi autonomamente: per tale motivo, le origini di lievito sono state chiamate **ARS** (*Autonomously Replicating Sequences*). Sono state individuate circa 350 ARS nel genoma di lievito, indicando che la dimensione media di un replicone in tali cellule è di circa 40 kpb (13,5 Mpb/350).

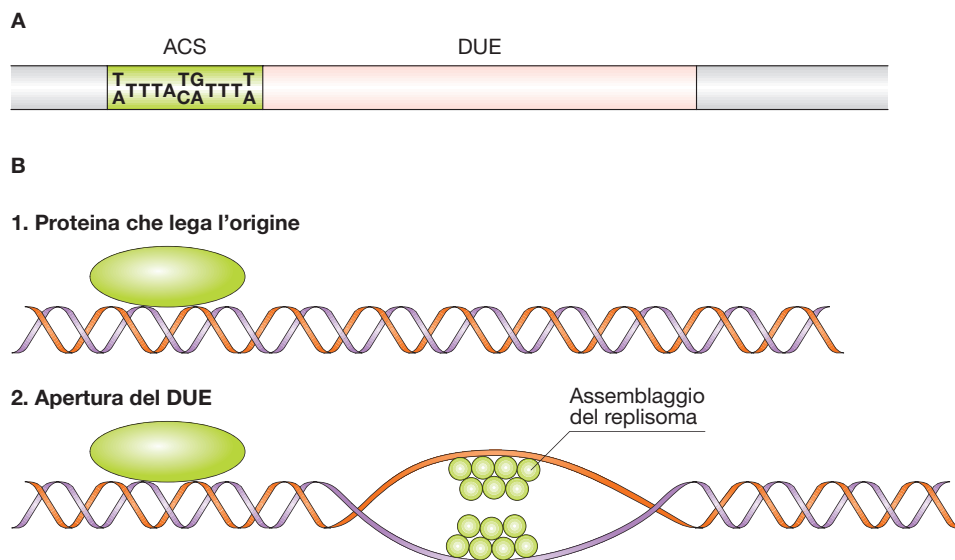


Figura 6.9 Struttura generale e attivazione di una ARS di lievito. (A) Elemento esemplificativo di tutte le ARS. L'ARS Consensus Sequence (ACS) è una sequenza di 11 pb conservata in tutte le ARS di lievito. L'ACS è essenziale per il funzionamento dell'origine ed

è fiancheggiata da almeno una regione ricca in A-T, più facilmente denaturabile, chiamata DUE (DNA Unwinding Element). (B) Il legame di una o più proteine iniziatrici porta all'apertura del DUE e al successivo assemblaggio del replisoma in tale regione.

Che caratteristiche ha una ARS di *S. cerevisiae*? Tutte le ARS di lievito contengono almeno una copia di una **ARS Consensus Sequence (ACS)**, talvolta chiamata anche **dominio A** di 11 pb [T/A]TTTA[T/C][G/A]TTT[T/A] (**fig. 6.9A**). Singole mutazioni all'interno dell'ACS aboliscono la funzione di una ARS. L'ACS è fiancheggiata da altre sequenze che hanno lunghezza variabile e la cui sequenza non è particolarmente conservata: sono comunque segmenti di DNA che, come l'ACS, sono ricchi in A e T. Si pensa, quindi, che siano regioni di DNA che possono essere facilmente aperte (definite come **DUE** o **DNA Unwinding Element**). La successione di eventi nell'utilizzo di un'origine sarebbe quindi il legame di proteine iniziatrici (*vedi* oltre) all'ACS, l'apertura del DUE e il successivo assemblaggio sull'origine della macchina di replicazione, chiamata anche **replisoma** (**fig. 6.9B**).

Origini precoci e origini tardive: influenza della struttura del cromosoma

Come già accennato precedentemente, non tutte le potenziali origini di replicazione sono effettivamente utilizzate in lievito e in altri eucarioti a ogni divisione cellulare e la loro accensione è sotto un preciso controllo temporale durante la fase S. L'accensione delle origini è geneticamente regolata sia da geni coinvolti

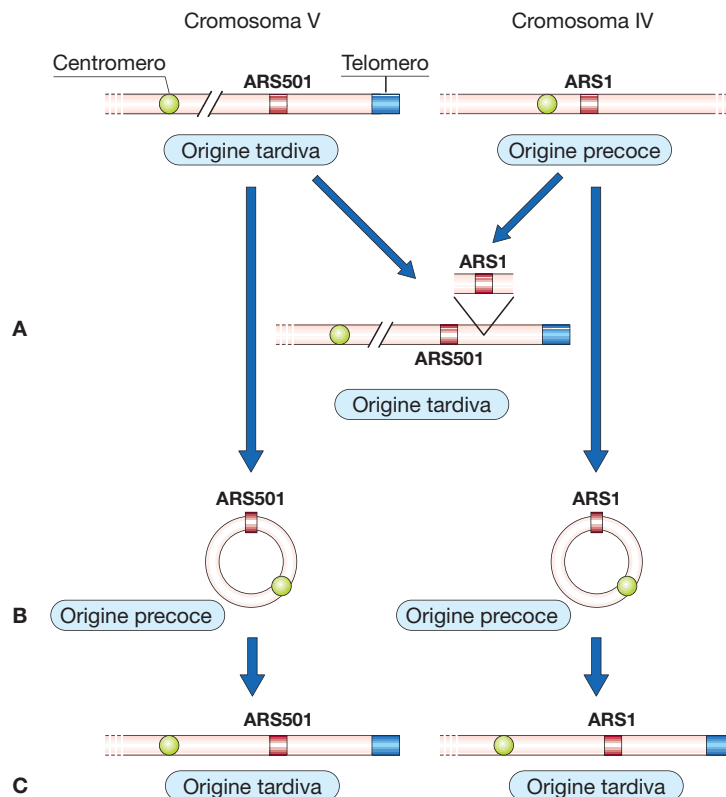


Figura 6.10 L'attivazione delle origini di replicazione in *Saccharomyces cerevisiae* è influenzata dal contesto cromosomico. Nello schema mostrato sono riassunti i risultati ottenuti spostando due ARS di lievito presenti su cromosomi diversi dal proprio contesto cromosomico a un altro. ARS1 si trova sul cromosoma IV vicino al centromero ed è accesa precocemente durante la fase S. ARS501 si trova sul cromosoma V vicino a una regione telomerica ed è attivata tardivamente in fase S. (A) È stato osservato che se, mediante tecniche di ricombinazione, ARS1 viene tolta dal suo contesto cromosomico normale e spostata vicino ad ARS501, nella regione

telomerica del cromosoma V, nella nuova posizione si accende tardivamente, indicando che il controllo temporale dell'accensione di un'ARS non dipende dalla sequenza di quella specifica ARS, ma dal contesto in cui si trova. (B) Analogamente, se entrambe le ARS sono spostate dalla loro localizzazione naturale su un plasmide, l'ARS501 che replicava tardivamente, induce una replicazione precoce del plasmide. (C) Al contrario, se i plasmidi vengono linearizzati e alle loro estremità sono aggiunte sequenze telomeriche, il nuovo contesto cromosomico determinato dalla vicinanza dei telomeri fa sì che i plasmidi linearizzati si replichino tardivamente.

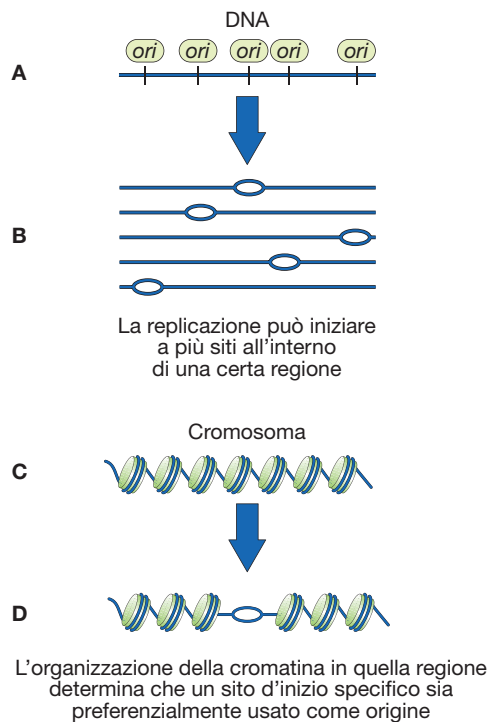


Figura 6.11 Struttura della cromatina e attivazione delle origini di replicazione nei vertebrati. Analogamente a quanto dimostrato in lievito, l'attivazione di un'origine di replicazione in eucarioti multicellulari è fortemente influenzata dalla struttura della cromatina. (A) In un certo segmento di DNA esistono molte potenziali origini (*ori*) e (B) tutte possono essere utilizzate, determinando una zona preferenziale d'attivazione della replicazione di quel segmento di DNA. Tuttavia, l'organizzazione del cromosoma in quella specifica regione (C) fa sì che una sola delle potenziali origini venga preferenzialmente utilizzata (D).

nel controllo della progressione del ciclo cellulare sia dal contesto cromosomico in cui si trovano. La struttura della cromatina influenza essenzialmente tutti i processi che riguardano il metabolismo del DNA, inclusa la replicazione. L'influenza del contesto cromosomico sull'attivazione delle origini è stata chiaramente dimostrata in lievito, dove è relativamente facile spostare sequenze di DNA dalla loro posizione originale a un'altra, ma è presente in tutti gli eucarioti. Si è visto, per esempio, che ARS1, un'origine precoce che si trova sul cromosoma IV vicino al centromero, diventa un'origine tardiva se spostata accanto ad ARS501, un'origine tardiva che si trova vicino al telomero sul cromosoma V (**fig. 6.10**). Lo stesso risultato si ottiene anche nell'esperimento inverso. Ciò dimostra in modo inequivocabile come il controllo temporale sull'accensione di un'origine non dipenda dalla sequenza nucleotidica dell'origine *per sé*, ma dal contesto cromosomico in cui essa si trova.

Il numero di origini di replicazione nell'uomo non è noto con precisione, ma si stima possano essere almeno 10^4 . Sia il numero che la struttura delle origini di replicazione nei mammiferi sono meno noti che in lievito perché, di fatto, le origini fino a ora identificate in mammiferi non sono in grado di rendere un elemento genetico che le possieda capace di replicarsi autonomamente se introdotto all'interno di una cellula. Di conseguenza, la loro identificazione e la definizione degli elementi strutturali importanti per la loro funzione sono meno definite. Come mostrato in **figura 6.11**, in una certa regione di DNA la replicazione può iniziare a siti diversi, identificando, eventualmente, una "zona" preferenziale d'inizio della replicazione. Quale sequenza funzioni davvero come origine dipende dalla struttura della cromatina, da altri aspetti del metabolismo del DNA, quale la trascrizione, e dal differenziamento cellulare.

Le origini di replicazione sui cromosomi possono essere mappate fisicamente basandosi sulla struttura insolita che le regioni di DNA assumono durante il processo replicativo. La tecnica dell'elettroforesi bidimensionale per studiare la struttura degli intermedi di replicazione (**finestra 6.1**) è stata ampiamente utilizzata

FINESTRA 6.1

MAPPATURA FISICA DELLE ORIGINI DI REPLICAZIONE

Nel testo è stato ricordato che lo screening genetico di una libreria di DNA ha portato all'identificazione di frammenti di DNA di lievito contenenti origini di replicazione: tali frammenti, infatti, conferiscono la capacità a un plasmide di *E. coli* di replicarsi autonomamente in cellule di lievito. Questo approccio non dimostrava, però, che fisicamente la replicazione iniziasse davvero a una specifica ARS nel suo contesto cromosomico. Inoltre, l'approccio genetico non ha permesso di isolare origini di replicazione da altri eucarioti.

Per risolvere questo problema è stata sviluppata una strategia di tipo biochimico. Tale approccio sperimentale si basa sull'ipotesi che un segmento di DNA che viene replicato debba assumere delle strutture caratteristiche (chiamate **intermedi di replicazione**) che dipendono dalla posizione relativa di quel segmento rispetto alla/e origine/i di replicazione da cui è partita l'ondata replicativa. Frammenti lineari di DNA di dimensioni (massa) diverse vengono facilmente separati uno dall'altro mediante elettroforesi (vedi cap. 20).

Un frammento di DNA in fase di replicazione non è lineare, ma possiede strutture che sono diverse a seconda della posizione dell'origine rispetto al frammento in esame. Tali strutture possono essere identificate mediante **elettroforesi bidimensionale del DNA**.

Prima di descrivere la tecnica utilizzata, cerchiamo di capire quali strutture possano formarsi sul DNA negli intermedi di replicazione. Nel pannello (B) della **figura 6F1.1** sono schematizzate alcune di tali strutture:

- I. struttura di un frammento che non contiene un'origine ma viene replicato da una forcella che entra nel frammento da destra. L'intermedio di replicazione che si forma ha una struttura a Y le cui braccia diventano via via più lunghe con il procedere della replicazione;
- II. il frammento non contiene un'origine ma viene invaso sia da destra che da sinistra da due forcelle di replicazione dando origine a due Y che si fonderanno all'interno del segmento stesso;

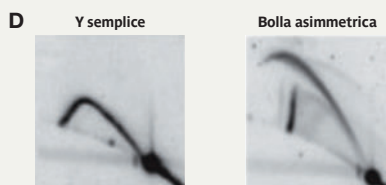
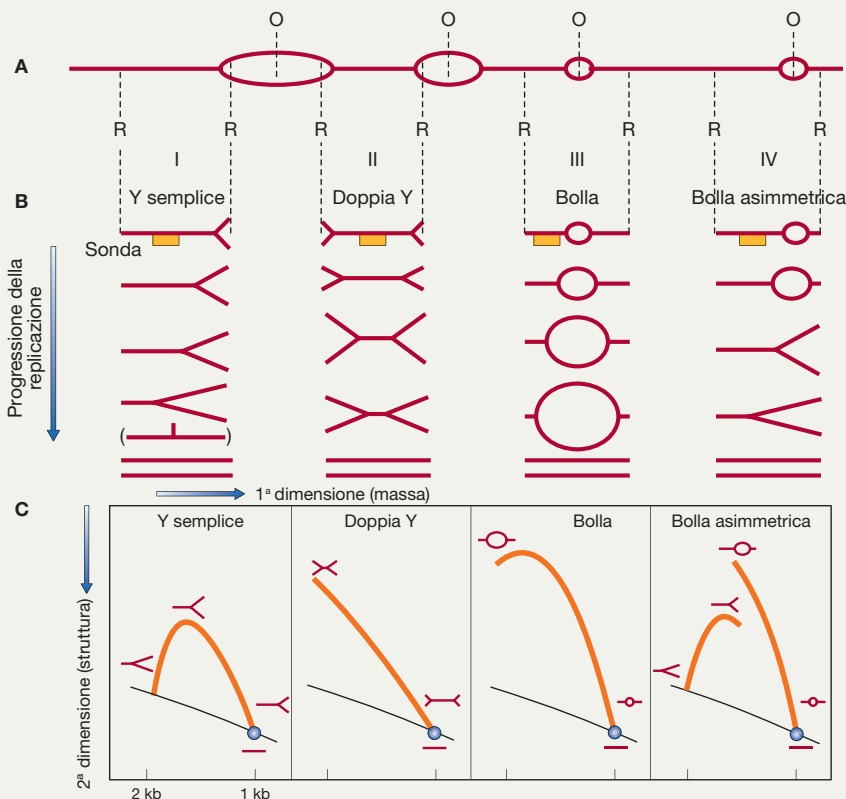


Figura 6F1.1 (A) Rappresentazione schematica del segmento cromosomico in esame contenente quattro origini di replicazione (O) e i siti di taglio per un enzima di restrizione (R). (B) Strutture che si originano, col progredire della replicazione, da ciascun frammento di restrizione e che dipendono dalla posizione delle origini relativamente ai frammenti stessi. (C) Schematizzazione dei risultati che si ottengono dopo separazione elettroforetica bidimensionale degli intermedi di replicazione e analisi di Southern con le sonde indicate in arancione in (B). (D) Due esempi di risultati sperimentali originali di elettroforesi bidimensionale.

- III. il frammento contiene un'origine di replicazione posta proprio al centro del segmento che si propaga bidirezionalmente: si formerà una bolla di replicazione che diventa via via più grande;
- IV. il frammento contiene un'origine di replicazione posta più vicina a una delle estremità: si formerà una bolla di replicazione che, espandendosi bidirezionalmente, finirà per originare una Y, riportandoci alla struttura considerata nel I caso.

Come accennato sopra, le strutture degli intermedi di replicazione possono essere analizzate e risolte mediante elettroforesi bidimensionale del DNA. Il DNA totale è isolato da cellule in divisione (quindi molte saranno in fase di attiva replicazione) e tagliato con uno specifico enzima di restrizione (A), così da generare una moltitudine di frammenti di DNA lineari o, eventualmente, contenenti intermedi di replicazione (B). I frammenti vengono poi separati mediante elettroforesi bidimensionale (C e D); nella 1^a dimensione (da sinistra a destra) la separazione avviene in base alla massa dei frammenti. Successivamente, il gel di elettroforesi viene ruotato di 90° e, cambiando le condizioni sperimentali, i frammenti di DNA sono separati nella 2^a dimensione (dall'alto in basso) in base alla loro geometria/struttura. Più la forma è geometricamente complessa, maggiore è l'ingombro sterico e, quindi, il ritardo elettroforético nella 2^a dimensione. Per poter visualizzare il frammento di DNA in esame si deve ricorrere alla tecnica di Southern (vedi cap. 20). Brevemente, dopo la separazione bidimensionale dei frammenti di DNA, questi vengono trasferiti su un

foglio di nitrocellulosa e la posizione delle diverse strutture del frammento in esame è identificata mediante ibridazione con la sonda indicata in (B) con un piccolo rettangolo arancione. Come schematizzato nel pannello (C), gli intermedi di replicazione visualizzati con la sonda danno origine a dei tracciati particolari, due dei quali sono mostrati anche in originale in (D). Per esempio, se il frammento è replicato da un'origine esterna come indicato in (I) si origina quello che è chiamato l'**arco delle Y**: via via che si allungano le braccia della Y il frammento sarà ritardato nella 1^a dimensione perché aumenta la sua massa, ma anche nella 2^a dimensione perché la presenza delle braccia aumenta l'ingombro sterico della molecola. Il massimo rallentamento nella 2^a dimensione si raggiunge quando le tre braccia della Y sono più o meno della stessa dimensione. Successivamente, il ritardo nella 2^a dimensione diminuisce perché (come indicato in BI tra parentesi) la molecola può essere considerata come un segmento di massa quasi raddoppiata, ma con un braccio molto corto. Se il frammento è replicato a partire da un'origine posizionata al centro del frammento (III) si forma una struttura con una bolla via via più grande che ritarda progressivamente sia nella 1^a che nella 2^a dimensione, originando il così detto "arco a bolla". Se il frammento è replicato a partire da un'origine interna ma asimmetrica nel frammento, si origina un iniziale "arco a bolla" che, nei frammenti in più avanzato stadio di replicazione, si tramuta nel tracciato atteso per un "arco delle Y". Questi ultimi tracciati sono i più informativi per determinare, mediante confronto con la sequenza nota del frammento in esame, la posizione di un'origine con un'approssimazione di pochi nucleotidi.

per analizzare in dettaglio la replicazione di elementi genetici extracromosomici, di cromosomi di lievito e anche di cromosomi di eucarioti multicellulari.

Oltre a permettere la mappatura delle origini di replicazione, tale tecnologia fornisce importanti informazioni sulla stabilità degli intermedi di replicazione in condizioni patologiche: quando, per esempio, il DNA è danneggiato per l'azione di agenti chimico-fisici (vedi cap. 7).

6.4 Meccanismo della replicazione del DNA nei procarioti

Sin dalla scoperta della struttura del DNA e della sua replicazione semiconservativa, si è cercato di capire come si realizzasse tale processo a livello molecolare, quali fossero le proteine coinvolte e come il meccanismo di replicazione fosse regolato così da avvenire una volta soltanto durante la fase S di ogni divisione cellulare. Come sempre accade in biologia, l'utilizzo di organismi modello è stato essenziale per comprendere i dettagli molecolari del meccanismo replicativo, anche se alcuni aspetti rimangono ancora oggi irrisolti.

Isolamento di mutanti letali-condizionali in *Escherichia coli* alterati nella replicazione del DNA

Informazioni importanti sono state ottenute utilizzando la replicazione di *E. coli* e dei suoi batteriofagi e plasmidi come modelli di replicazione nei procarioti. Il problema è stato affrontato in numerosi laboratori combinando approcci speri-

mentali di tipo genetico e di tipo biochimico. L'approccio genetico è stato molto importante perché ha permesso di isolare mutanti difettivi nella replicazione del DNA. Tale processo è essenziale per la vita di una cellula: pertanto, mutazioni nella replicazione causano normalmente la morte della cellula. Mutazioni in geni essenziali per la vitalità cellulare devono, quindi, ricadere nella categoria delle **mutazioni letali-condizionali**. Tali mutazioni si esprimono in una certa condizione sperimentale (detta restrittiva), ma non in un'altra (detta permissiva): mutanti letali-condizionali in un processo essenziale possono essere così tenuti in vita nella condizione permissiva, mentre si può valutare l'effetto della mutazione dopo trasferimento del mutante alla condizione restrittiva. La categoria di mutazioni letali-condizionali più utilizzata in laboratorio è quella dei **mutanti temperatura-sensibili (ts)**. Nel caso delle cellule di *E. coli*, un ceppo selvatico (wild-type) cresce indifferentemente a 37 o a 42 °C, mentre un mutante ts cresce a 37 °C (temperatura permissiva), ma non a 42 °C (temperatura restrittiva) (**finestra 6.2**). Questo accade perché la mutazione determina un cambiamento amminoacidico che altera la struttura tridimensionale della proteina codificata dal gene mutato, così che non è più funzionale a 42 °C, mentre mantiene la sua attività a 37 °C.

L'isolamento di mutanti ts ha portato all'individuazione di decine di geni di *E. coli* codificanti per proteine che svolgono la loro funzione nella replicazione del DNA. L'identificazione delle proteine codificate da tali geni ha tenuto impegnati i ricercatori fino a pochi anni fa e ancora oggi alcuni aspetti della loro funzione nel processo replicativo non è chiarita.

“Replisoma” di *Escherichia coli*

Diversi fattori hanno contribuito a rendere difficile l'identificazione e caratterizzazione biochimica delle diverse proteine richieste per la replicazione del DNA. Innanzitutto, come nella maggior parte dei meccanismi coinvolti nel metabolismo del DNA, la sua replicazione non è portata avanti dall'azione di proteine che

FINESTRA 6.2

ISOLAMENTO DI MUTAZIONI TEMPERATURA-SENSIBILI IN *ESCHERICHIA COLI*

In (A) è descritta la procedura per l'isolamento di mutazioni temperatura-sensibili che prende il nome di **replica-plating**.

Una coltura batterica viene fatta crescere per diverse generazioni in un terreno di crescita liquido, dopo essere stata sottoposta al trattamento con agenti mutageni (raggi UV, raggi X, sostanze chimiche mutagene), così da aumentare la frequenza di mutazione.

Aliquote della coltura batterica mutagenizzata vengono piastrate su numerose “piastre madri”, così che ogni cellula batterica dia origine a una colonia osservabile individualmente sulle piastre.

Un cilindro coperto con un velluto sterile viene delicatamente appoggiato su ciascuna delle piastre madri, così che sul velluto rimanga un'impronta delle colonie presenti su una specifica piastra.

Le cellule presenti sul velluto vengono depositate su due piastre: una viene incubata alla temperatura permissiva (37 °C), l'altra alla temperatura restrittiva (42 °C).

Dopo 1-2 giorni di incubazione si paragonano le due piastre derivate da ciascuna piastra madre e incubate alle diverse temperature per individuare le colonie che crescono a 37 °C, ma non a 42 °C. Ogni colonia si è originata da una cellula che ha subito una mutazione temperatura-sensibile. Parago-

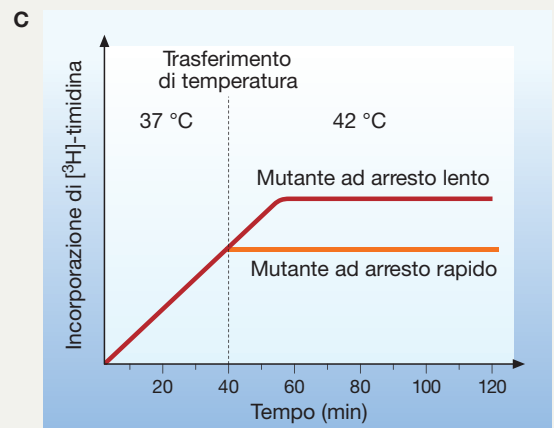
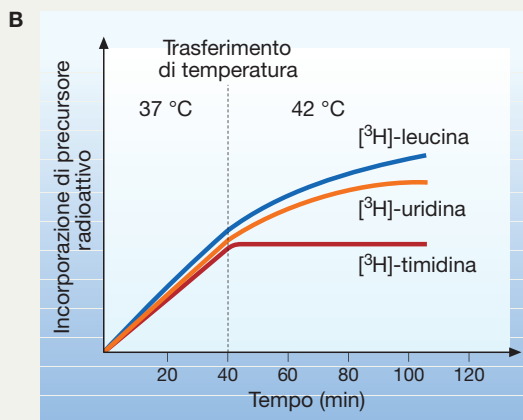
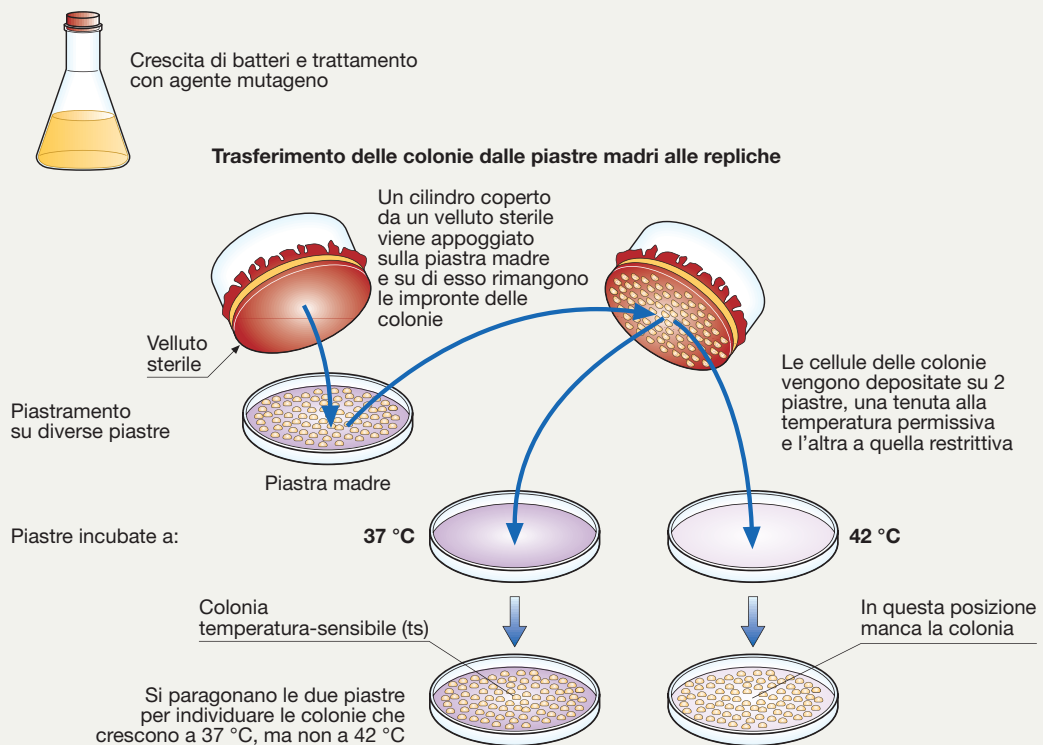
nando la posizione della colonia nelle piastre figlie e nella piastra madre, si può ritornare a quest'ultima per prelevare la colonia temperatura-sensibile originaria e procedere alla caratterizzazione funzionale della mutazione (B e C).

(B) Avendo isolato un mutante temperatura-sensibile, è ora importante capire quale processo molecolare sia alterato dalla mutazione. Limitandoci alle biosintesi macromolecolari più importanti nella cellula, la mutazione potrebbe bloccare una reazione richiesta per la sintesi del DNA o dell'RNA o delle proteine. Per discriminare tra queste tre possibilità, colture indipendenti di cellule mutate (e per controllo cellule wild-type) vengono fatte crescere alla temperatura permissiva di 37 °C in presenza di diversi precursori radioattivi: [³H]-timidina per analizzare la sintesi di DNA, [³H]-uridina per valutare la sintesi di RNA e [³H]-leucina per misurare la sintesi delle proteine. Dopo 40 minuti a 37 °C le cellule sono trasferite alla temperatura restrittiva di 42 °C, sempre in presenza di uno o dell'altro dei precursori radioattivi. Ai tempi indicati, si prelevano aliquote delle cellule. Le cellule vengono lisate e la macromolecola radioattiva marcata nei tre esperimenti indipendenti è raccolta su filtri che, posti in un apparecchio in grado di rilevare la radioattività, ci danno una misurazione della sintesi di DNA,

RNA o proteine (in ordinata nel grafico in **B**). Nell'esempio mostrato nel grafico in (**B**) risulta evidente che l'incorporazione nel DNA di $[^3\text{H}]$ -timidina si blocca rapidamente dopo trasferimento della coltura a 42°C , mentre la sintesi di RNA e proteine continua per un certo tempo. Questo risultato permette di stabilire che la mutazione in esame altera la sintesi del DNA. Se il blocco rapido avesse riguardato la sintesi di RNA o proteine, avremmo concluso che la mutazione alterava una o l'altra di queste due biosintesi macromolecolari. Una volta determinato che la mutazione altera la sintesi del DNA, possiamo valutare l'incorporazione di $[^3\text{H}]$ -timidina

nel DNA in condizioni più controllate e analizzare, nel corso dell'esperimento, campioni a tempi di crescita più ravvicinati. Come indicato nel grafico in (**C**), questo esperimento, la cui base sperimentale è identica a quella descritta in (**B**), permette di stabilire se la mutazione porta a un arresto rapido o a un arresto più lento nella sintesi del DNA. Come è descritto più in dettaglio nel testo, nel primo caso la mutazione affligge una funzione richiesta per l'allungamento della sintesi del DNA, mentre un blocco lento identifica una mutazione che altera la funzione di una proteina coinvolta nell'inizio della replicazione del DNA.

A Isolamento di mutanti temperatura-sensibili



agiscono indipendentemente una dall'altra. Al contrario, le proteine replicative funzionano come pezzi di un'unica "macchina" proteica (chiamata **replisoma**) che richiede il perfetto assemblaggio e le corrette interazioni chimiche fra tutti i componenti.

Macchine proteiche agiscono essenzialmente in tutti i meccanismi coinvolti nel controllo dell'espressione genica, come la trascrizione, la ricombinazione, la traduzione ecc. La stabilità di queste macchine proteiche può essere, però, molto diversa: per esempio, il ribosoma, la macchina proteica coinvolta nella traduzione, è così stabile da poter essere isolato essenzialmente nella sua integrità. Al contrario, il replisoma è molto instabile e allorché si rompono le cellule di *E. coli* o di qualunque altro organismo, i vari componenti del replisoma si dissociano uno dall'altro e, per assemblare nuovamente la macchina, è necessario recuperare i vari pezzi e rimetterli insieme. Ciò è reso molto complicato dal fatto che la concentrazione delle proteine che costituiscono il replisoma è molto bassa all'interno della cellula, per cui è molto difficile identificarne i componenti e purificarli dall'insieme di tutte le altre proteine cellulari. L'identificazione, la purificazione e la ricostituzione del replisoma hanno tenuto impegnati i ricercatori per decine d'anni.

Scoperta della DNA polimerasi e sue proprietà biochimiche

L'approccio sperimentale strettamente biochimico ed enzimologico è stato molto importante nella fase iniziale della ricerca. Infatti, il modello della replicazione semiconservativa del DNA suggerisce immediatamente l'esistenza di enzimi in grado di catalizzare la polimerizzazione di nucleotidi. Se durante la replicazione semiconservativa del DNA ciascuno dei due filamenti parentali serve da stampo per la neosintesi di un filamento complementare, c'è da aspettarsi che nella cellula esistano delle attività in grado di catalizzare la formazione del legame fosfodiesterico tra nucleotidi in modo DNA stampo-dipendente. Tale ipotesi portò all'identificazione della prima **DNA polimerasi** in cellule di *E. coli* da parte di Arthur Kornberg (**fig. 6.12**), il cui laboratorio ha dato i contributi più importanti nel campo dell'enzimologia della replicazione del DNA. Numerose DNA polimerasi sono state successivamente trovate sia in *E. coli* che nelle cellule eucariotiche (**finestra 6.3 e tabella 6.3**).

La sintesi del DNA sia *in vivo* (cioè nella cellula) che *in vitro* (cioè in provetta tramite l'utilizzo di componenti purificati) ha delle precise richieste biochimiche. I substrati della reazione sono i deossiribonucleosidi trifosfati (dATP, dGTP, dCTP e dTTP, chiamati in generale **dNTP**) (**fig. 6.13A**), ma un secondo substrato richiesto da tutte le DNA polimerasi sino a oggi conosciute è anche un **complesso innesco-stampo** (*primer-template*) (**fig. 6.13B**) costituito, appunto, da uno stampo rappresentato da un filamento di DNA e da un innesco che può essere un tratto più o meno lungo di DNA (o, come vedremo, spesso l'innesco può essere di RNA) che porti un'estremità 3'OH che la DNA polimerasi sia in grado di allungare. Infatti, tutte le DNA polimerasi, a differenza delle RNA polimerasi, non sono in grado di iniziare la sintesi di una catena nucleotidica, ma necessitano di un'estremità 3'OH da allungare. La DNA polimerasi aggiunge, a ogni evento di condensazione, un nuovo nucleotide, catalizzando la formazione del legame fosfodiesterico nel rispetto della regola della complementarietà tra le basi con il filamento di DNA stampo.

Nella formazione di ciascun legame fosfodiesterico (**fig. 6.13C**), il fosfato in posizione α del dNTP viene legato al 3'OH dell'innesco portando alla liberazione di pirofosfato. L'energia libera di questa reazione è piuttosto modesta ($\Delta G = -3,5$ kcal/mole), e l'energia aggiuntiva che spinge la reazione verso la polimerizzazione è fornita dall'idrolisi del pirofosfato da parte di un enzima chiamato **pirofosfatasi**.

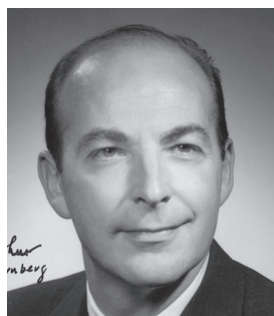


Figura 6.12 Arthur Kornberg. Ha per primo identificato, isolato e analizzato una DNA polimerasi. (Fonte: <http://profiles.nlm.nih.gov/WH/B/B/M/G/>.)

FINESTRA 6.3

IN *ESCHERICHIA COLI* NON TUTTE LE DNA POLIMERASI SONO COINVOLTE NELLA REPLICAZIONE DEL DNA

In *Escherichia coli* sono presenti 5 DNA polimerasi, chiamate I, II, III, IV e V.

La DNA polimerasi I è stata la prima DNA polimerasi a essere purificata nel laboratorio di Arthur Kornberg, e per qualche tempo si pensò fosse l'enzima coinvolto nella replicazione del DNA del batterio. Per questo motivo, ad Arthur Kornberg fu assegnato il premio Nobel nel 1959. In realtà più tardi si scoprì che la DNA polimerasi I è soprattutto coinvolta nella riparazione del DNA e nella replicazione svolge un ruolo nella maturazione dei frammenti di Okazaki (vedi testo). Kornberg è stato, comunque, lo scienziato che successivamente ha scoperto la vera DNA polimerasi replicativa di *E. coli* e nel suo laboratorio sono state identificate e purificate numerose altre

proteine replicative, per cui l'assegnazione del premio Nobel è stata più che meritata, anche se la motivazione originale fu sbagliata!

La DNA polimerasi II è anch'essa coinvolta in meccanismi di riparazione del DNA.

La **DNA polimerasi III** è il vero enzima replicativo in *E. coli*. L'oloenzima è composto da 10 subunità e si assembla in tappe successive (vedi fig. 6.19 e testo) formando un dimero in grado di sintetizzare entrambi i nuovi filamenti di DNA.

La DNA polimerasi IV (chiamata anche DinB) e la DNA polimerasi V sono due polimerasi translesione (vedi testo) in grado di aggirare danni al DNA che bloccano, almeno temporaneamente, la macchina replicativa.

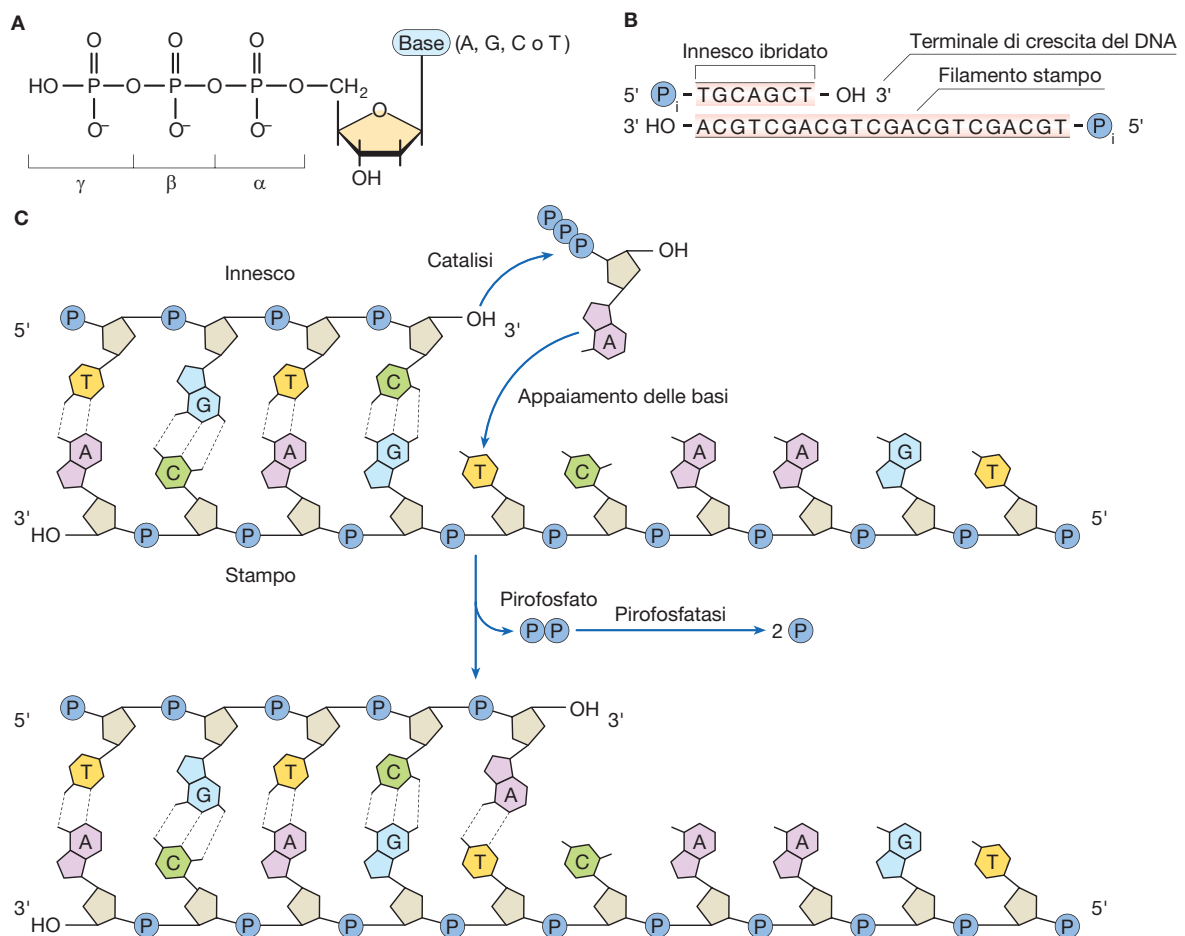


Figura 6.13 Substrati necessari e meccanismo di sintesi del DNA da parte di una DNA polimerasi. (A) È indicata la struttura generale di un deossiribonucleoside trifosfato: la posizione dei tre fosfati relativamente alla molecola dello zucchero è indicata con le lettere dell'alfabeto greco. (B) È indicata la struttura del complesso innesco-stampo. L'innesco deve avere un terminale OH libero all'estremità 3'. Si forma, quindi, una regione a doppio filamento tra lo stampo e l'innesco, mentre la regione dello stampo che verrà copiata dalla DNA polimerasi mediante l'allungamento dell'estremità 3' OH

dell'innesco è a singolo filamento. Nel caso della replicazione del DNA *in vivo*, l'innesco fisiologico è normalmente rappresentato da una corta molecola di RNA (vedi testo). (C) L'allungamento dell'innesco e la successiva sintesi di tutto il DNA prevedono l'attacco nucleofilo del fosfato α del dNTP che deve essere aggiunto all'estremità 3' OH. Ciò determina l'allungamento di un nucleotide della catena nascente di DNA e la liberazione di una molecola di pirofosfato che, essendo rapidamente degradata da una pirofosfatasi, fornisce gran parte dell'energia necessaria per la biosintesi del DNA.

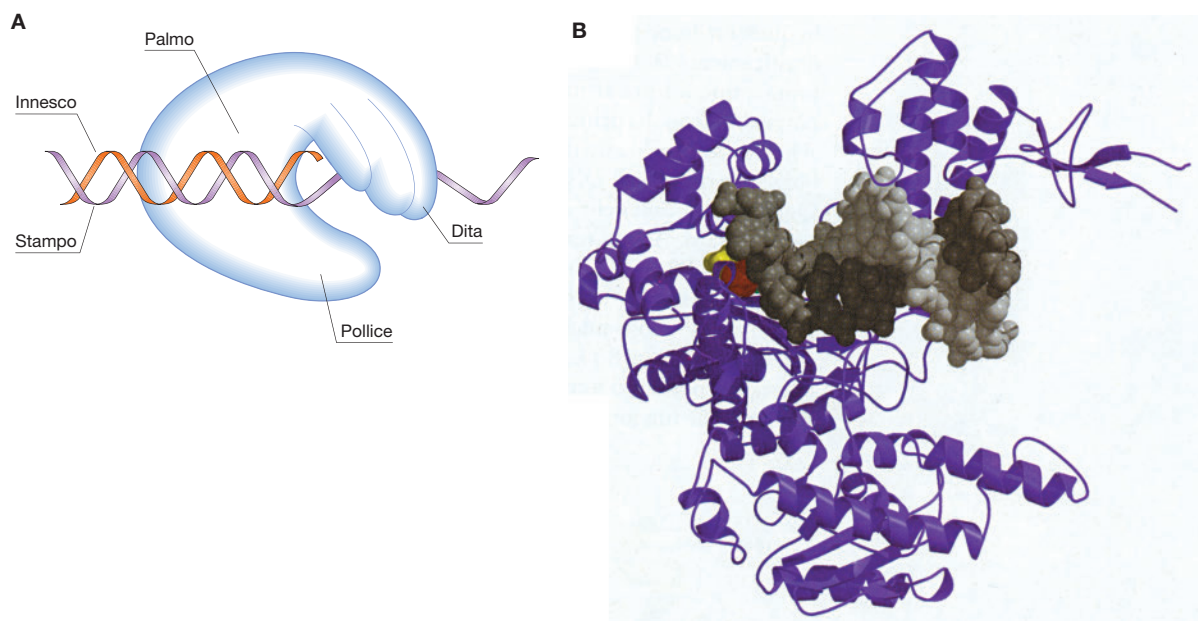


Figura 6.14 Struttura tridimensionale di una DNA polimerasi replicativa. (A) La struttura tridimensionale di una DNA polimerasi replicativa determinata dall'analisi cristallografica del complesso innesco-stampo-enzima fa pensare a una mano destra in cui si distinguono palmo, dita e pollice. La funzione delle diverse

parti della mano è descritta nel testo. (B) Struttura del complesso enzima-DNA della polimerasi del batteriofago T7. Il DNA e i tre domini principali (palmo, dita e pollice) della proteina sono facilmente identificabili. [Fonte: Doublié, S. et al., 1998, *Nature*, 391, p. 251.]

Da questa breve descrizione della reazione si evince una caratteristica importante della sintesi del DNA: la sua polarità procede in direzione $5' \rightarrow 3'$ e tutte le DNA polimerasi funzionano seguendo tale polarità.

Una spiegazione molecolare di come la DNA polimerasi riesca a catalizzare la polimerizzazione del DNA è venuta dalla definizione della struttura tridimensionale di complessi innesco-stampo-enzima. La struttura delle DNA polimerasi replicative è stata paragonata a quella di una mano destra parzialmente chiusa dove si distinguono tre domini proteici definiti, per analogia, pollice, dita e palmo (**fig. 6.14**). Il complesso innesco-stampo è principalmente associato con il palmo, mentre il sito catalitico della DNA polimerasi si trova tra le dita e il pollice. La presenza di ioni bivalenti (di norma Mg^{++} o Zn^{++}) è importante per favorire la polimerizzazione dei dNTP. Il palmo influenza fortemente l'accuratezza della sintesi del DNA. Infatti, appaiamenti tra stampo e innesco che non rispettano la complementarità delle basi abbassano la velocità della reazione, e il complesso innesco-stampo è più facilmente rilasciato dal sito attivo della polimerizzazione entrando in contatto con il sito esonucleasico dell'enzima che controlla la fedeltà del processo replicativo (*vedi oltre*). Le dita dell'enzima svolgono anch'esse dei ruoli importanti per favorire la fedeltà della replicazione; il pollice dell'enzima, invece, non è direttamente coinvolto nella catalisi enzimatica, ma interagisce con lo stampo stabilizzando il complesso ternario innesco-stampo-enzima e favorendo un'azione processiva dell'enzima stesso che si dissocia dallo stampo solo dopo aver catalizzato il legame di numerosi nucleotidi.

Fedeltà di replicazione del DNA

La fedeltà del processo replicativo è impressionante: circa 1 errore ogni 10^9 - 10^{10} nucleotidi polimerizzati. Un sistema basato solo sulla stereochimica di coppie di basi che rispettino la regola della complementarità non sarebbe in grado di raggiungere l'accuratezza indicata sopra. La selettività delle polimerasi è piuttosto li-

mitata (1 errore ogni circa 10^5) per la presenza di forme tautomeriche delle basi azotate che determinano inevitabilmente degli errori che devono essere eliminati (**fig. 6.15A**). Questi errori sono rimossi da un'attività esonucleasica $3' \rightarrow 5'$ che ha, quindi, una polarità inversa alla direzione di sintesi del DNA. Tale attività esonucleasica è associata alle polimerasi replicative vere e proprie (ma non a tutte le DNA polimerasi) e ha la funzione di agire come **correttore di bozze** (attività *proofreading*) (**fig. 6.15B**). Quando la DNA polimerasi rileva la presenza di un appaiamento non corretto tra le coppie di basi, il complesso stampo-innesco si allontana dal sito catalitico di polimerizzazione della DNA polimerasi e si avvicina al sito esonucleasico con funzione *proofreading*. Tale attività elimina il nucleotide errato e permette alla polimerasi di riprendere la sintesi senza che il complesso ternario innesco-stampo-enzima si sia dissociato. In media, le DNA polimerasi replicative fanno un errore ogni circa 10^5 nucleotidi polimerizzati. L'attività di correttore di bozze della polimerasi riduce la frequenza di errore a uno ogni 10^7 nucleotidi. La frequenza di errore (e quindi l'insorgenza di mutazioni) è ulteriormente diminuita tramite uno specifico sistema di riparazione del DNA (*vedi cap. 7*) che riduce ulteriormente il numero di appaiamenti non corretti durante il processo replicativo.

Alcune DNA polimerasi procariotiche presentano anche un dominio esonucleasico con polarità $5' \rightarrow 3'$ (**fig. 6.15B**): questa attività ha una funzione rilevante in DNA polimerasi coinvolte in processi di riparazione del DNA e svolge, in alcuni casi, una funzione importante nella rimozione degli inneschi di RNA utilizzati nel processo replicativo (*vedi oltre*). L'accoppiamento tra attività della DNA polimerasi e azione della esonucleasi $5' \rightarrow 3'$ è alla base di una reazione che viene chiamata **nick translation**, tecnica di laboratorio che permette di sintetizzare frammenti di DNA marcati con precursori radioattivi (**fig. 6.15C**) da usarsi come sonde radioattive in vari tipi di esperimenti (*vedi finestra 20.1*).

La DNA polimerasi, che è la parte della macchina replicativa richiesta per la formazione del legame fosfodiesterico tra i nucleotidi, agisce, da quanto detto so-

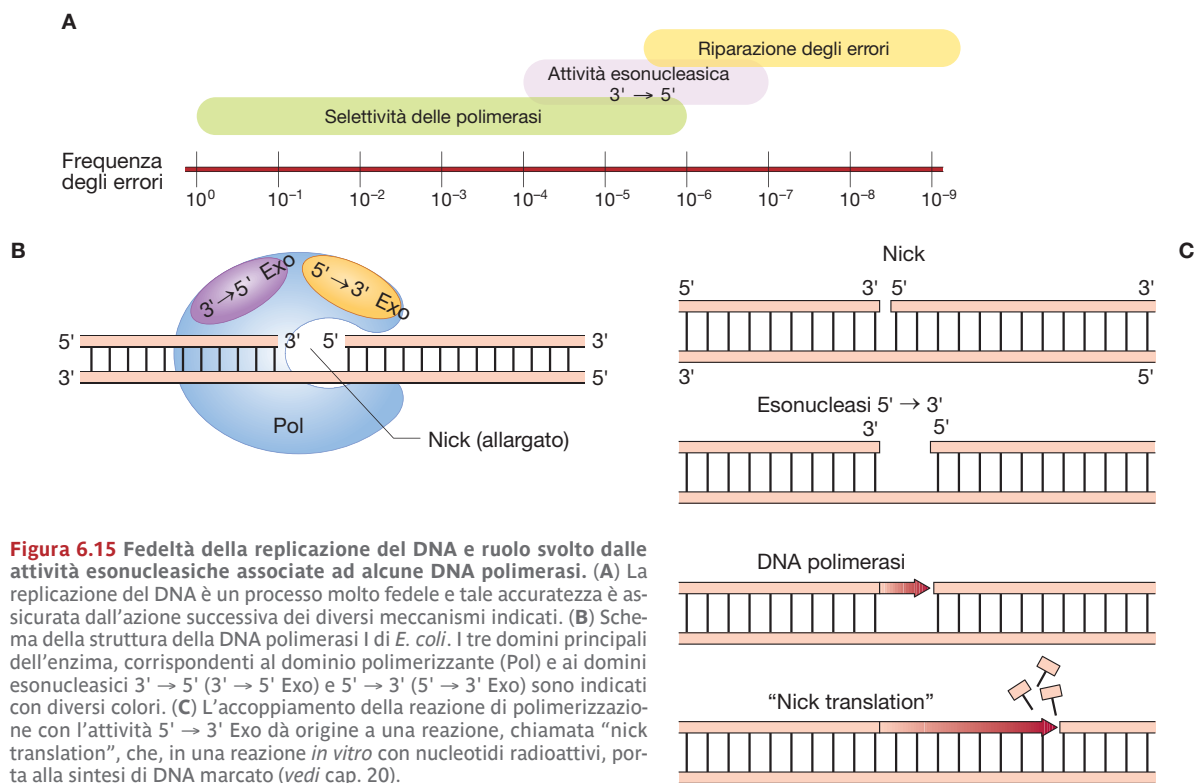


Figura 6.15 Fedeltà della replicazione del DNA e ruolo svolto dalle attività esonucleasiche associate ad alcune DNA polimerasi. (A) La replicazione del DNA è un processo molto fedele e tale accuratezza è assicurata dall'azione successiva dei diversi meccanismi indicati. (B) Schema della struttura della DNA polimerasi I di *E. coli*. I tre domini principali dell'enzima, corrispondenti al dominio polimerizzante (Pol) e ai domini esonucleasici $3' \rightarrow 5'$ ($3' \rightarrow 5'$ Exo) e $5' \rightarrow 3'$ ($5' \rightarrow 3'$ Exo) sono indicati con diversi colori. (C) L'accoppiamento della reazione di polimerizzazione con l'attività $5' \rightarrow 3'$ Exo dà origine a una reazione, chiamata "nick translation", che, in una reazione *in vitro* con nucleotidi radioattivi, porta alla sintesi di DNA marcato (*vedi cap. 20*).

pra, utilizzando *in vitro* un substrato costituito da un complesso stampo-innesco. È chiaro che *in vivo* lo stampo è fornito dai due filamenti della molecola di DNA parentale, ma chi fornisce l'innesco e come viene coordinata la sintesi dei due filamenti di nuova formazione?

Nella cellula entrambi i filamenti di neosintesi vengono sintetizzati contemporaneamente e, perché ciò avvenga, è necessaria la separazione dei due filamenti della molecola di DNA parentale. La regione che si trova tra i due filamenti separati sui quali avviene la neosintesi del DNA e la doppia elica che non ha ancora subito il processo di apertura, come abbiamo già visto, viene chiamata **forcella di replicazione** (*replication fork*) (**fig. 6.16A**).

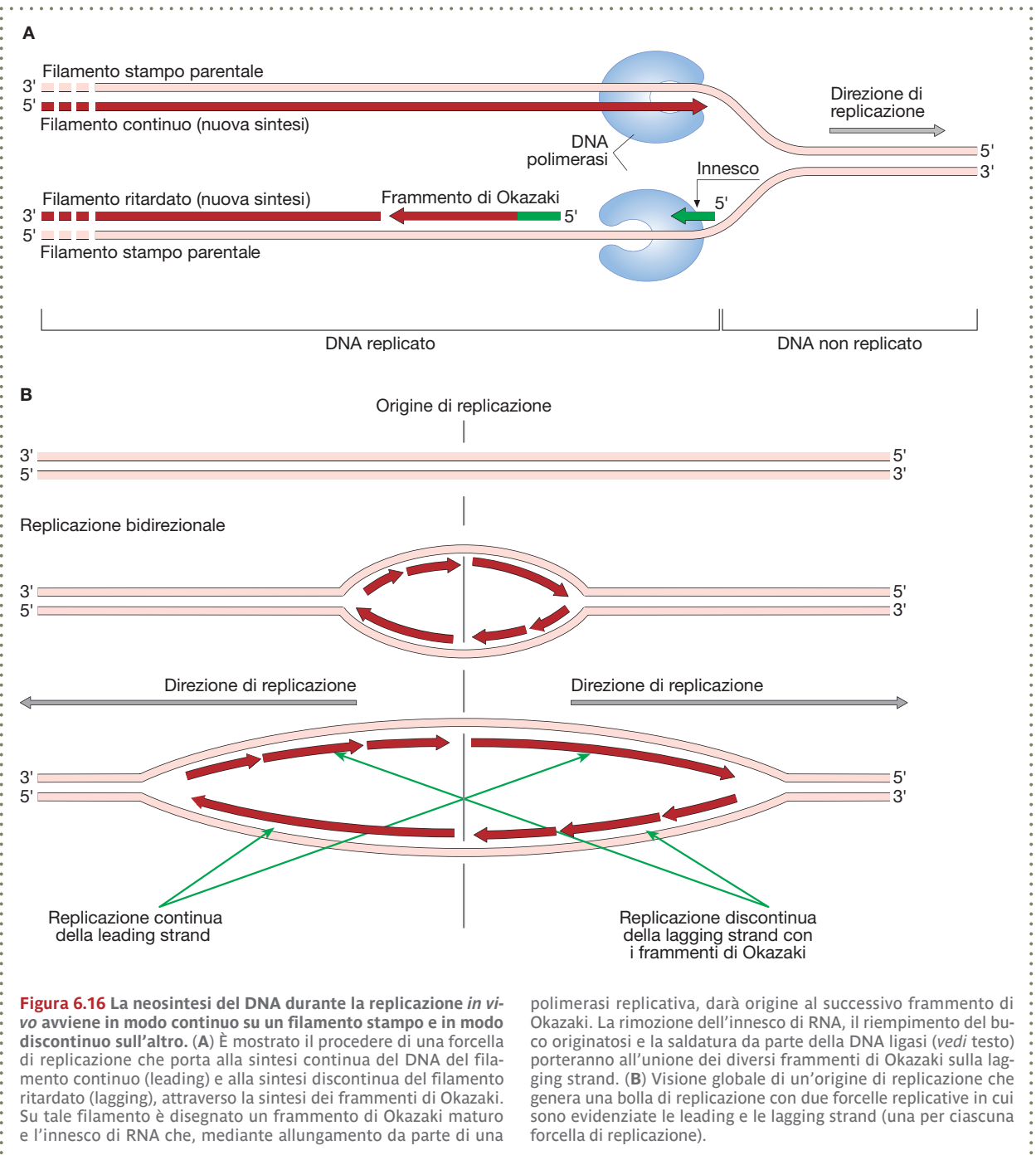
Forcella di replicazione: sintesi del filamento continuo e del filamento discontinuo

È importante avere una visione della forcella replicativa nel contesto dell'accensione di un'origine di replicazione che procede bidirezionalmente (**fig. 6.16B**). In tale contesto, la forcella di replicazione mostrata in **figura 6.16A** altro non è che una delle due metà di una bolla di replicazione con l'adiacente DNA non ancora replicato. Tenendo presente che tutte le DNA polimerasi sono in grado di polimerizzare il DNA soltanto in direzione $5' \rightarrow 3'$, ciò crea un problema nella progressione della forcella replicativa. Infatti, come mostrato in **figura 6.16A e B**, soltanto uno dei due filamenti di neosintesi può essere sintetizzato in modo continuo seguendo la direzione della forcella replicativa: questo filamento di neosintesi viene chiamato **filamento continuo** o *leading strand*. Per rispettare la polarità di sintesi del DNA ($5' \rightarrow 3'$) l'altro filamento deve essere sintetizzato in modo discontinuo e genera quello che viene definito **filamento ritardato** o *lagging strand*. Tale filamento è definito "ritardato" perché la sua sintesi può iniziare solo dopo che il progredire della forcella di replicazione ha generato sufficiente DNA a singolo filamento da far partire la sintesi del filamento "lagging". La sintesi del filamento ritardato si realizza attraverso la generazione di frammenti discontinui di DNA che vengono chiamati **frammenti di Okazaki** dal nome del loro scopritore: i frammenti di Okazaki in *E. coli* sono lunghi 1000-2000 nucleotidi mentre negli eucarioti sono molto più corti (100-200 nucleotidi). Il coordinamento tra la sintesi continua della leading strand e la sintesi discontinua della lagging strand, così come il processamento e la saldatura dei frammenti di Okazaki tra di loro, sono tra i problemi più complessi che la macchina replicativa deve riuscire a risolvere.

Innesco della sintesi di DNA

L'innesco della sintesi di ogni frammento di Okazaki, così come l'innesco del filamento leading a un'origine di replicazione, devono prevedere la sintesi di un iniziatore (*primer*) per offrire alla DNA polimerasi il complesso stampo-iniziatore descritto poco sopra. L'innesco della sintesi di DNA sia sulla leading che sulla lagging strand è fornito da corte (4-12 nucleotidi) molecole di RNA sintetizzate da un enzima, denominato **DNA primasi**. L'utilizzo di molecole di RNA "primer" come inneschi della sintesi del DNA implica che la macchina replicativa riesca poi a eliminare tali tratti di RNA e sia in grado di riempire i buchi (*gaps*) che vengono così generati.

Questi semplici esempi ci ricollegano a quanto descritto poco sopra sulla complessità della macchina replicativa (replisoma) e sul fatto che l'analisi genetica aveva individuato decine di geni codificanti per proteine che svolgono la loro funzione nella replicazione del DNA. Abbiamo accennato che per alcune di queste proteine è stato possibile immaginare una funzione e, quindi, lo sviluppo di saggi biochimici che ne permettessero la purificazione e caratterizzazione. Abbiamo fatto l'esempio della DNA polimerasi e un altro caso è stato rappresentato dall'enzima **DNA ligasi**, la cui esistenza era stata ipotizzata dopo aver osservato



che la lagging strand veniva sintetizzata in modo discontinuo, creando, quindi, la necessità di un enzima in grado di saldare il 3'OH di un frammento di Okazaki con il 5'-fosfato del frammento di Okazaki sintetizzato precedentemente.

Come sono state identificate, quindi, tutte le altre proteine oggi note che costituiscono il replisoma di *E. coli*? Nei procarioti ciò è stato possibile utilizzando dei sistemi ricostituiti *in vitro* in cui si cercavano tutte le proteine richieste per replicare molecole semplici di DNA a struttura nota. Uno dei sistemi che si sono rilevati più utili per identificare proteine coinvolte nella fase d'inizio della sintesi del DNA e nel suo allungamento è stato lo studio della replicazione *in vitro* del DNA circolare a singolo filamento di alcuni batteriofagi, quali G4, M13 e Φ X174. Questi

Tabella 6.1 Elenco di alcuni geni e funzioni coinvolti nella replicazione del DNA in *Escherichia coli*.

Gene	Fenotipo del mutante ts	Funzione della proteina
<i>DnaA</i>	Arresto ritardato	Riconoscimento di <i>oriC</i>
<i>DnaB</i>	Arresto rapido/ritardato	DNA elicasi
<i>DnaC</i>	Arresto rapido/ritardato	Stimola la DNA elicasi
<i>DnaG</i>	Arresto rapido	DNA primasi
<i>DnaE</i>	Arresto rapido	Subunità α della DNA polimerasi III
<i>DnaN</i>	Arresto rapido	Subunità β della DNA polimerasi III
<i>DnaQ</i>	Mutatore	Subunità ϵ della DNA polimerasi III
<i>Ssb</i>	Arresto rapido, difetti di riparazione e ricombinazione	Proteina che lega il DNA a singolo filamento (ssDNA)
<i>Lig</i>	Accumulo di frammenti Okazaki	DNA ligasi

La terminologia *DnaA*, *DnaB*, *DnaC* ecc. è spesso usata per indicare le proteine codificate dai rispettivi geni. Le proteine sono alternativamente indicate in base alle loro funzioni biochimiche: DNA primasi, DNA polimerasi, SSB ecc.

batteriofagi iniettano il loro DNA circolare a singolo filamento (*single-stranded* o ssDNA) nel batterio ospite e questo viene trasformato in una molecola di DNA circolare a doppio filamento (*double-stranded* o dsDNA) chiamata **forma replicativa (RF)** del DNA del batteriofago. La replicazione di ssDNA a formare DNA a doppio filamento ha richieste enzimatiche più semplici che non la replicazione di classiche molecole di DNA a doppia elica, per cui l'utilizzo di tali modelli ha permesso di identificare più facilmente un certo numero di proteine replicative. L'identificazione delle proteine che provvedono al riconoscimento e all'apertura dell'origine di replicazione di *E. coli* e alla duplicazione del suo DNA circolare a doppia elica ha richiesto, invece, l'utilizzo di vettori plasmidici che contenessero, appunto, l'*oriC* del batterio ed è stata molto più complessa.

Saggi di complementazione *in vitro*

L'approccio sperimentale utilizzato per l'individuazione e purificazione delle proteine richieste per la replicazione del DNA delle molecole sopra citate è stato geniale e dimostra quanto possa essere produttivo combinare la logica sperimentale della genetica con quella della biochimica. Tale approccio è stato utilizzato in molti laboratori ma, ancora una volta, il laboratorio pilota nell'individuare la strategia sperimentale è stato quello di Arthur Kornberg. Senza entrare in eccessivi dettagli, il presupposto è molto semplice. Il genoma dei batteriofagi di *E. coli* con ssDNA circolare si replica utilizzando, essenzialmente, funzioni codificate nel genoma del batterio ospite: quindi, i batteriofagi non si replicano in condizioni restrittive (42 °C) nei mutanti ts di *E. coli* alterati nella replicazione del DNA (**finestra 6.2**), mentre si replicano a temperatura permissiva (37 °C).

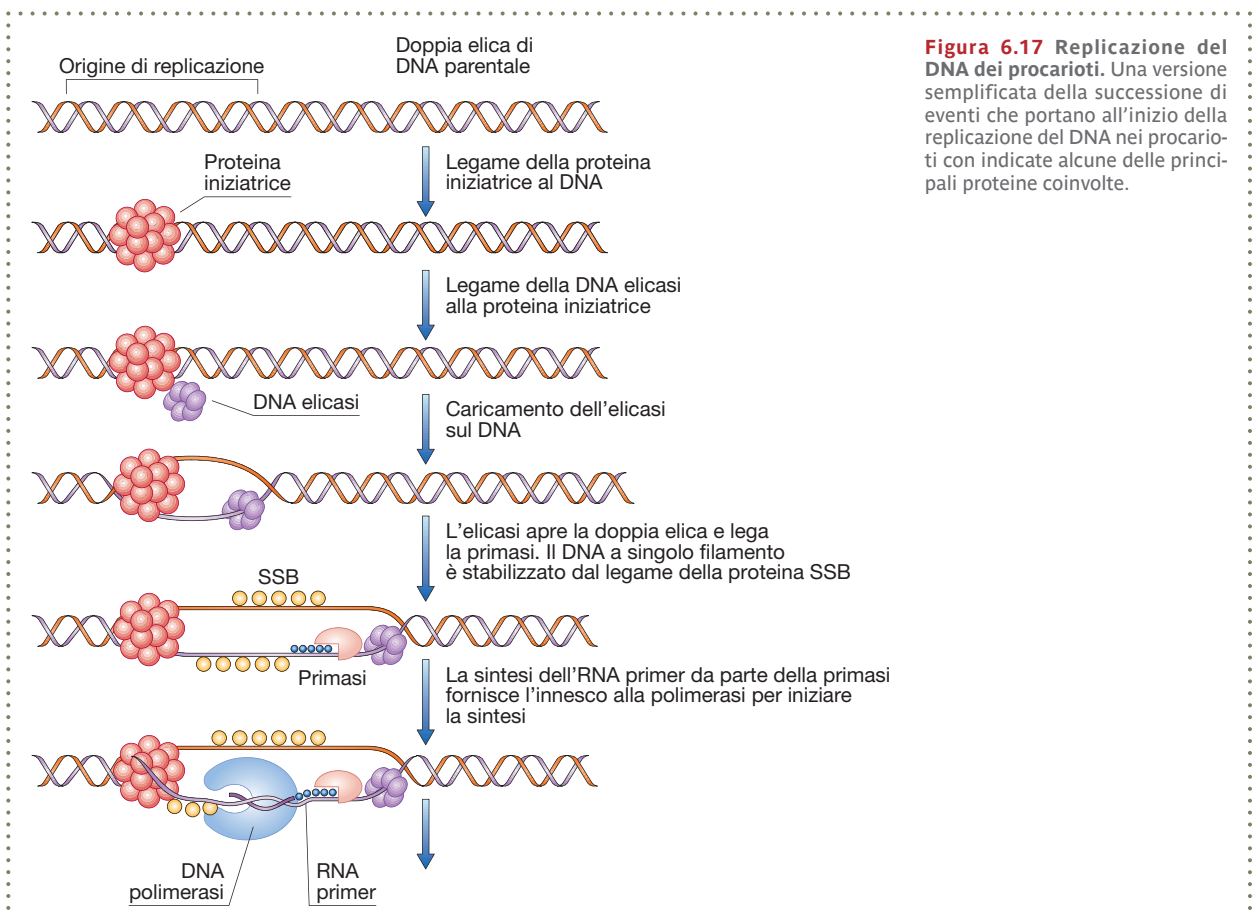
Quello che si osserva *in vivo* si verifica anche *in vitro*: infatti, se a un estratto proteico di cellule mutanti ts di *E. coli* si aggiunge l'ssDNA circolare di un batteriofago, le proteine presenti nell'estratto sono in grado di convertire l'ssDNA in molecole di dsDNA, ammesso che l'estratto sia incubato con il DNA stampo alla temperatura permissiva. Questo non si verifica, invece, alla temperatura restrittiva, dato che la proteina mutata richiesta per la replicazione non è funzionale a 42 °C. Se però all'estratto preparato dal mutante si aggiunge un'aliquota dell'estratto preparato dal ceppo di *E. coli* selvatico, la proteina normale presente in tale aliquota è in grado di sostituire l'attività che è alterata nell'estratto del mutante (**finestra 6.4**). Tale esperimento prende il nome di **complementazione *in vitro*** e rispecchia quanto succede *in vivo*, dove l'espressione del gene selvatico è normalmente in grado di complementare l'effetto di una mutazione nello stesso gene.

Quindi, anche se non si conosceva a livello biochimico la funzione molecolare codificata dal gene mutato, è stato possibile purificare la proteina codificata da quel gene andando a isolare, attraverso diversi passaggi di purificazione, la proteina in questione dall'estratto del ceppo selvatico, seguendo la capacità di questa proteina di complementare il difetto della sua forma mutata. In questo modo sono state purificate le proteine codificate da molti geni coinvolti nella replicazione del DNA la cui funzione era ignota. La successiva caratterizzazione biochimica, strutturale e funzionale delle proteine in questione ha permesso di capire il loro ruolo nel processo replicativo. Un elenco parziale di geni e proteine coinvolti nella replicazione del DNA di *E. coli* è identificati con strategie analoghe a quelle sopra descritte è indicato in **tabella 6.1**.

Proteine replicative di *Escherichia coli*

Un quadro complessivo, anche se estremamente semplificato, di quanto succede sul DNA durante la replicazione, nei procarioti in genere, è rappresentato dallo schema mostrato in **figura 6.17**. Nella figura sono indicate semplicemente le funzioni delle principali proteine coinvolte senza entrare in alcun dettaglio molecolare. Un quadro più dettagliato della successione di eventi molecolari che portano a iniziare la replicazione del cromosoma e delle proteine coinvolte è riportato qui di seguito per il caso specifico di *E. coli*.

1. Il primo evento molecolare consiste nel riconoscimento di *oriC* da parte della proteina **DnaA**. Come anticipato descrivendo la struttura di *oriC*, l'origine minima della replicazione di *E. coli* contiene 4 ripetizioni di 9 pb e 3 ripetizioni di 13 pb (**fig. 6.18A**). Come mostrato in **figura 6.18B**, DnaA si attacca



inizialmente alle ripetizioni di 9 pb: la proteina si lega in modo cooperativo formando una specie di nucleo centrale proteico intorno al quale si avvolge il DNA di *oriC*. Successivamente, DnaA si lega anche alle 3 ripetizioni di 13 pb facilitando la denaturazione localizzata del DNA in quella regione e permettendo l'assemblaggio di altre proteine replicative.

2. Il passaggio successivo consiste nel caricamento di **DnaB** e **DnaC** a livello della bolla di denaturazione creatasi nella regione di *oriC* dando origine a un complesso proteico, denominato **complesso di pre-innesco** della sintesi del DNA, che può essere visualizzato anche al microscopio elettronico. DnaB possiede un'attività elicasica, per cui consumando ATP è in grado di separare i due filamenti di DNA. Lo svolgimento del DNA sia nella fase iniziale che nel successivo processo di allungamento della sintesi del DNA genera una tensione torsionale del DNA che è risolta da enzimi in grado di modificare la topologia del DNA e chiamati "DNA topoisomerasi" (vedi cap. 2).
3. La bolla di denaturazione creatasi nella regione dell'origine tende a rinaturare: per evitare ciò, il DNA a singolo filamento originatosi per azione di DnaA e dell'elicasa DnaB-DnaC viene stabilizzato dal legame della **proteina SSB** alla regione di ssDNA (**fig. 6.17**). SSB (*Single-Stranded DNA Binding protein*) è

FINESTRA 6.4

PURIFICAZIONE DI UNA PROTEINA E SAGGIO DI COMPLEMENTAZIONE *IN VITRO*

In questo testo abbiamo spesso parlato della funzione essenziale di singole proteine o di macchine proteiche e la Biologia Molecolare ha spesso lo scopo di identificare tutte le proteine coinvolte in un certo processo biologico e di ricostituire questo processo in provetta (*in vitro*) utilizzando proteine purificate. La struttura, la funzione e la purificazione di proteine sono descritte in testi di Biochimica, a cui rimandiamo. Vale però la pena di ricordare cosa significa purificare una proteina e come i saggi di complementazione *in vitro* hanno portato a purificare un numero consistente di proteine coinvolte nella replicazione del DNA senza che la loro esatta funzione biochimica fosse conosciuta.

Purificare una proteina significa adottare una serie di procedure successive per cui alla fine la proteina in esame è isolata da tutte le altre proteine della cellula ed è possibile studiare in dettaglio la sua struttura e funzione.

Le diverse proteine cellulari hanno proprietà diverse in termini di struttura, dimensioni, carica, interazioni con substrati ecc. I biochimici utilizzano queste diverse proprietà per cercare di isolare una proteina da tutte le altre, proprio sfruttando le peculiari caratteristiche della proteina in esame.

Nella tabella mostrata nel pannello (A) un estratto cellulare è stato purificato utilizzando strategie successive che permettono di separare le proteine in base alla loro carica (cromatografia su fosfocellulosa o DEAE-cellulosa), dimensione (separazione su sefariosio), interazioni con substrati (legame a una resina di cellulosa su cui è stato legato del DNA) ecc. Nel caso di una proteina con attività enzimatica è possibile seguire la proteina in esame nei vari passaggi di purificazione misurando la sua **attività enzimatica**, cioè la capacità di catalizzare una determinata reazione. Facendo l'esempio di una DNA polimerasi, la sua attività enzimatica sarà, per esempio, la capacità di catalizzare il legame fosfodiesterico tra dNTP copiando un DNA stampo in presenza di un opportuno innesco. L'attività di un enzima viene espressa in unità enzimatiche: l'unità di un enzima è una misura arbitraria stabilita

dai ricercatori ed è espressa, continuando l'esempio della DNA polimerasi, in nmoli di dNTP polimerizzati in una certa unità di tempo. Il successo della purificazione di una proteina con attività enzimatica viene valutato quantificando il miglioramento dell'attività specifica dell'enzima nei diversi passaggi di purificazione. L'attività specifica è il rapporto tra unità enzimatiche e quantità di proteine totali: se tale rapporto aumenta nei diversi passaggi di purificazione significa che riusciamo a mantenere (o perdiamo relativamente poco) l'attività enzimatica, mentre via via scartiamo le proteine a cui non siamo interessati.

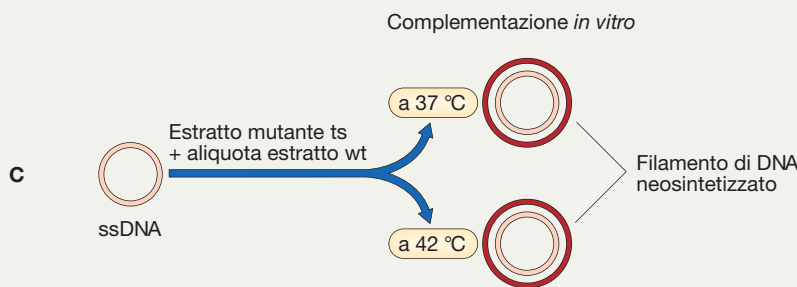
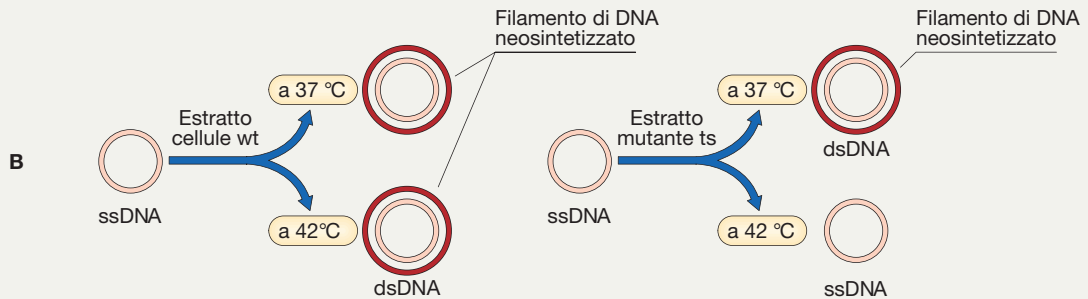
Nell'esempio indicato la proteina è stata purificata 10 000 volte nei 5 passaggi (a-e) indicati: l'attività specifica, infatti, passa da 5 unità/mg a 50 000 unità/mg con un recupero dell'attività enzimatica iniziale del 10% (da 500 000 unità a 50 000 unità).

I **saggi di complementazione *in vitro*** sono stati sviluppati per purificare proteine replicative di cui non si conosceva l'attività enzimatica: la logica dell'approccio sperimentale è descritta nel testo. Brevemente, come riassunto in (B), un estratto di *E. coli* wild-type è in grado di replicare un DNA circolare a singolo filamento (ssDNA) in uno a doppio filamento (dsDNA) sia a 37 che a 42 °C. L'estratto preparato da un mutante temperatura-sensibile (ts) è in grado di svolgere la stessa reazione solo alla temperatura permissiva di 37 °C, ma non a 42 °C. Se, però (C), all'estratto del mutante ts si aggiunge un'aliquota dell'estratto wild-type, la proteina selvatica presente in quest'ultimo estratto è in grado di complementare il difetto della proteina mutata, per cui si osserverà replicazione anche alla temperatura restrittiva di 42 °C.

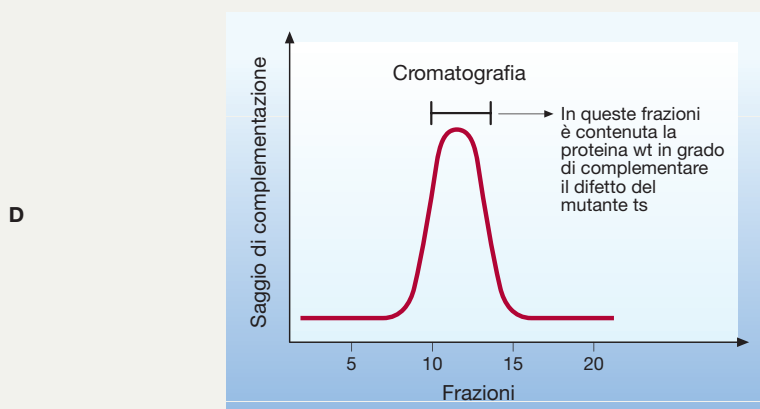
Sarà ora possibile (D) cercare di purificare la proteina in questione dall'estratto wild-type, seguendo nei vari passaggi di purificazione (qui è esemplificato un singolo ipotetico passaggio cromatografico) l'attività capace di complementare il difetto di replicazione a 42 °C dell'estratto preparato dal mutante ts.

Schema arbitrario della purificazione di una proteina

Frazione	Proteine totali (mg)	Attività enzimatica (unità)	Attività specifica (unità/mg)	Purificazione
a) Estratto	100 000	5×10^5	5	–
b) Fosfocellulosa	4000	4×10^5	100	20 volte
c) DEAE-cellulosa	200	2×10^5	1000	200 volte
d) Sefarosio	20	1×10^5	5000	1000 volte
e) DNA-cellulosa	1	5×10^4	50 000	10 000 volte

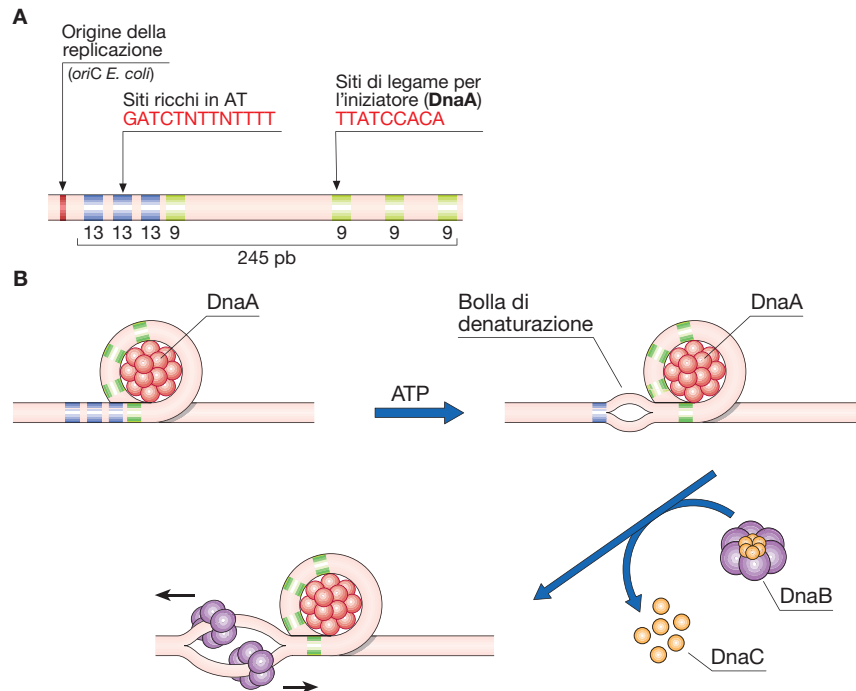


Purificazione dal ceppo wild-type della proteina in grado di complementare *in vitro*



una proteina che ha un'affinità maggiore per il DNA a singolo filamento che a doppio filamento. Gioca, quindi, un ruolo rilevante in tutte le reazioni del metabolismo del DNA in cui è prevista una produzione transiente di ssDNA. Anche SSB, come la proteina DnaA, interagisce con il DNA in modo cooperativo: questo termine sta a indicare che il legame tra una molecola di SSB all'ssDNA facilita il legame di una seconda molecola di SSB alla stessa molecola di DNA.

Figura 6.18 Inizio della replicazione del DNA in *E. coli*. (A) Schema dell'*oriC* di *E. coli*: la più corta sequenza di DNA in grado di funzionare come origine di replicazione è costituita da 245 pb e include 3 ripetizioni di 13 pb e quattro ripetizioni di 9 pb indicate con diversi colori. (B) Numerosi monomeri della proteina DnaA si legano, inizialmente, alle 4 ripetizioni di 9 pb. I monomeri DnaA, agendo in modo cooperativo, formano un complesso proteico intorno al quale si arrotola il DNA. Ciò determina, in presenza di ATP, una parziale apertura della doppia elica nella regione delle tre ripetizioni di 13 pb. Tale bolla di denaturazione rappresenta il sito d'entrata dell'elicasi replicativa costituita da un esamero di DnaB. In presenza della proteina accessoria DnaC, due complessi esamerici di DnaB vengono caricati sulla bolla di denaturazione, definendo, così, le due forcelle di replicazione bidirezionale.



4. Come già più volte detto, tutte le DNA polimerasi necessitano di un'estremità 3'OH pre-formata per poter aggiungere i nucleotidi successivi. In alcuni casi il 3'OH è generato dall'introduzione di un taglio in una molecola di DNA a doppio filamento oppure è addirittura fornito da una proteina (**finestra 6.5**). In *E. coli* e negli eucarioti, però, gli inneschi che forniscono il 3'OH che può essere allungato dalla DNA polimerasi sono corte molecole di RNA; tali molecole funzionano da primer sia per far partire la sintesi continua a un'origine del filamento leading sia per iniziare la sintesi di tutti i frammenti di Okazaki sul filamento ritardato. La **DNA primasi di *E. coli***, codificata dal gene *dnaG*, è costituita da un singolo polipeptide di circa 60 kDa e la sintesi degli **RNA primer** costituisce il primo effettivo evento di sintesi nella replicazione del DNA.
5. La DNA polimerasi replicativa di *E. coli* è la **DNA polimerasi III oloenzima**, una macchina proteica molto complessa formata da numerose proteine che si organizzano in una serie di sub-complessi a formare un dimero, se non addirittura, un trimero catalitico. Il "sub-complesso catalitico" vero e proprio è formato da 3 subunità chiamate α (alfa) (attività DNA polimerizzante), ϵ (epsilon) (attività esonucleolitica correttore di bozze 3' $\rightarrow$ 5') e θ (teta) (attività di stimolazione della esonucleasi). Ciascun "sub-complesso catalitico", però, può assemblarsi sul DNA solo tramite l'azione del "sub-complesso γ (gamma)". Negli ultimi anni, in virtù di una serie di dati strutturali e biochimici, è risultato sempre più evidente che il "sub-complesso γ " è il principale organizzatore della struttura dell'intero oloenzima sul DNA. Da che proteine è composto il "sub-complesso γ " e qual è la sua funzione? Il "sub-complesso γ " è composto da 3 copie della subunità γ , 1 copia ciascuna delle subunità δ (delta), δ' (delta primo), χ (chi) e ψ (psi) e 3 copie della subunità τ (tau). Da un punto di vista enzimatico il "sub-complesso γ " appartiene alla famiglia delle proteine AAA⁺ che utilizzano l'energia di legame e idrolisi dell'ATP per modificare la struttura di proteine bersaglio e/o del DNA. Da un punto di vista biochimico possono essere classificate come delle ATPasi DNA-dipendenti. L'idrolisi dell'ATP è utilizzata per aprire e caricare sul DNA il "sub-complesso β ": tale sub-complesso è costituito da 2 copie della subunità β che formano un omo-

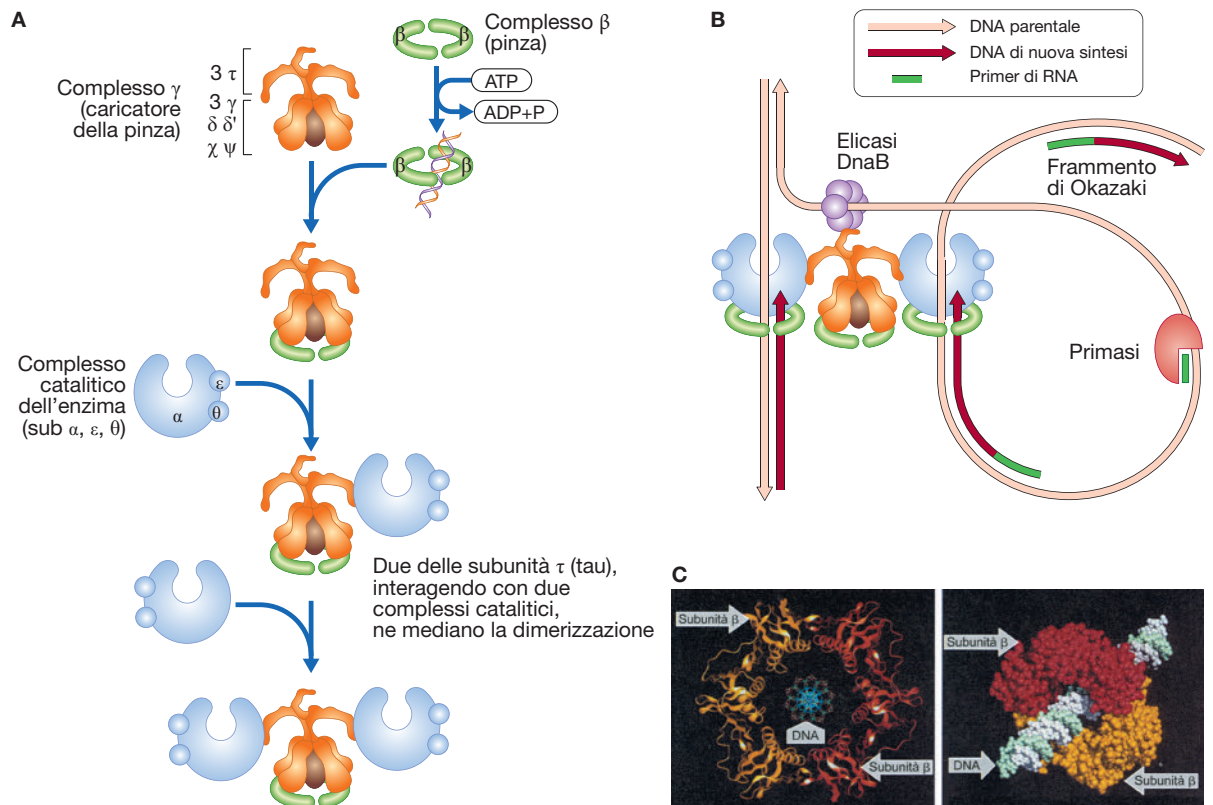


Figura 6.19 Assemblaggio della DNA polimerasi III oloenzima sul DNA. (A) La DNA polimerasi III oloenzima è formata da numerose subunità proteiche assemblate in 3 distinti sub-complessi: complesso γ , complesso β e complesso catalitico (vedi testo). Il complesso γ è formato da proteine diverse (3γ , 3τ , 1δ , $1\delta'$, 1ψ , 1χ) e svolge la funzione di caricare il complesso β e coordinare la funzionalità dei 2 complessi catalitici tramite le subunità τ . (B) L'elicasi DnaB apre il DNA e interagisce con una subunità τ , mentre le altre 2 subunità τ connettono i 2 complessi catalitici tenuti

sul DNA dall'interazione con il complesso β . Il complesso catalitico che sintetizza il filamento continuo rimane associato al DNA, mentre il complesso che sintetizza discontinuamente il filamento lagging si dissocia alla fine della sintesi di ciascun frammento di Okazaki, per poi riassociarsi all'innesco sintetizzato dalla primasi nella regione dell'ansa. Questo modello porta il filamento ritardato ad agire parallelamente a quello continuo ed è chiamato "modello a trombone", per la sua somiglianza con lo strumento musicale. (C) Modelli strutturali del complesso β .

dimero con una struttura a ciambella, chiamato anche "pinza β " (" β -clamp"), in grado di abbracciare la molecola di DNA che passa nel foro della ciambella (figg. 6.19A e 6.19C). La "pinza β " ha la funzione di tenere il "sub-complesso catalitico" sul DNA e facilitarne lo scorrimento. Qual è la funzione delle subunità τ presenti nel "sub-complesso γ "? Tali subunità non giocano alcun ruolo nel caricamento della "pinza β " dato che l'attività ATPasica DNA-dipendente non richiede la presenza delle subunità τ . Il ruolo di τ è, invece, di contattare (in virtù della loro flessibilità strutturale) i due dimeri catalitici così che essi siano in grado di replicare in modo coordinato il filamento continuo (leading) e il filamento ritardato (lagging) della doppia elica. Questo concetto è solo apparentemente in contraddizione con il fatto che, avendo i due filamenti polarità opposte, i due nuclei catalitici dovrebbero anch'essi muoversi in direzioni opposte. Infatti, il coordinamento tra sintesi del filamento continuo e sintesi del filamento ritardato è mantenuto come mostrato in figura 6.19B. L'aspetto più importante del modello (che manca ancora di alcune evidenze sperimentali) è che la sintesi del filamento leading su uno dei due filamenti stampo induce la formazione, sul filamento, lagging, di un'ansa che fornisce lo stampo per la sintesi del filamento ritardato. Il "sub-complesso γ ", che serve per caricare la "pinza β ", rimane associato al nucleo catalitico che agisce sul filamento ritardato. Il "sub-complesso γ " serve per caricare ma anche per

scaricare dal DNA la “pinza β ” e questo può permettere il riciclo del “sub-complesso catalitico” dell’oloenzima richiesto per l’innescio di ogni frammento di Okazaki sul filamento ritardato.

6. Nel DNA completamente replicato non possono venire inglobati i piccoli frammenti di RNA sintetizzati dalla DNA primasi: questi vengono principalmente rimossi dall’attività esonucleolitica 5'→3' della DNA polimerasi I che, mentre elimina gli RNA primer, è in grado di riempire i buchi o “gap” che essa stessa genera (*vedi* descrizione della reazione di nick translation, **fig. 6.15**). I “nick”, cioè le interruzioni che rimangono tra i vari frammenti di Okazaki, sono, infine, saldati dall’azione della DNA ligasi.

Quanto sopra descritto è una semplificazione dell’intero processo di sintesi del DNA in *E. coli*, ma offre una visione della successione degli eventi richiesti nel processo stesso.

Dobbiamo ricordare, inoltre, che il genoma di *E. coli* è costituito da un unico replicone: la sintesi di DNA inizia a *oriC* formando, così, due forcelle di replicazione che procedono bidirezionalmente. Come termina la replicazione quando le due forcelle replicative si incontrano? La **terminazione della replicazione** di *E. coli* è controllata da ripetizioni di una breve sequenza di 23 pb, chiamata **ter**. Come indicato in **figura 6.20**, le sequenze **ter** sono posizionate a circa 180° rispetto a *oriC* e le diverse sequenze fanno terminare l’una o l’altra delle due forcelle di replicazione. Per impedire il movimento delle forcelle di replicazione le sequenze **ter** devono legare una proteina chiamata **TBP** (*Ter Binding Protein*) o **Tus**. Durante la terminazione della replicazione è anche necessaria l’azione di DNA topoisomerasi (*vedi* cap. 2) per separare fisicamente le due molecole di DNA circolari che si sono originate.

FINESTRA 6.5

MODALITÀ DI GENERAZIONE DELL'ESTREMITÀ 3'OH RICHIESTA COME INNESCO DALLE DNA POLIMERASI

Pannello A. Vengono descritti i modi più comuni per generare l’estremità 3'OH richiesta dalle DNA polimerasi per la sintesi di DNA.

- I. DNA a doppio filamento o a singolo filamento senza inneschi con estremità 3'OH non possono essere copiati da alcuna delle DNA polimerasi note.
- II. Nella maggior parte degli organismi l’estremità 3'OH è costituita da un RNA primer generato dall’enzima DNA primasi.
- III. In alcuni casi il 3'OH è generato da un taglio endonucleolitico che provoca un’incisione o “nick”, seguito da una reazione chiamata “strand displacement” o “spostamento del filamento”.
- IV. Nella replicazione di alcuni virus il 3'OH è fornito da un nucleotide legato a una proteina che agisce essa stessa da innesco.

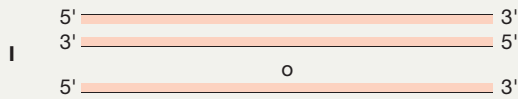
Pannello B. La produzione di molecole di DNA di alcuni batteriofagi di *E. coli* avviene nel batterio tramite il meccanismo del cerchio rotante. Come già descritto nel testo, alcuni batteriofagi di *E. coli*, quali ϕ X174, G4, M13, contengono un DNA a singolo filamento circolare (ssDNA). Quando tali batteriofagi infettano una cellula batterica, iniettano al suo interno il loro ssDNA e questo viene convertito in un DNA a doppio filamento circolare, chiamato anche “forma replicativa” (RF)

del DNA del batteriofago. La replicazione dell’ssDNA avviene, essenzialmente, mediante la sintesi di un RNA primer sintetizzato dalla primasi del batterio e suo successivo allungamento e saldatura a formare il DNA a doppia elica della RF. L’RF del batteriofago, attraverso il meccanismo del cerchio rotante, porta alla sintesi di molte molecole di ssDNA che verranno poi assemblate in capsidi generando nuovi fagi infettanti. L’ssDNA originario è indicato con il segno + e genera l’RF indicata come +/- . Un taglio sul filamento + genera un’estremità 3'OH che venendo via via allungata sposta l’estremità 5' del filamento +, generando una specie di coda (“strand displacement”). Procedendo nella reazione, si genera una coda molto lunga che mediante taglio endonucleolitico forma molte unità + dell’ssDNA del fago che, una volta circolarizzate e inglobate nei capsidi, generano nuove particelle fagiche mature.

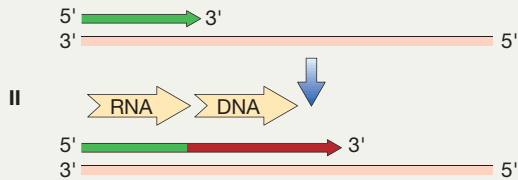
Pannello C. In alcuni virus (per esempio l’adenovirus che infetta cellule animali) il 3'OH richiesto per la replicazione è fornito da un nucleotide covalentemente attaccato a una proteina. La proteina è chiamata “proteina terminale” (TP) perché si attacca all’estremità del DNA a doppio filamento del virus: un suo residuo di serina può formare un legame covalente con il dCTP, generando un complesso TP-dCMP che fornisce il 3'OH richiesto per la replicazione del DNA operata dalla DNA polimerasi del virus che interagisce con la proteina TP e sintetizza il DNA mediante il meccanismo dello “strand displacement”.

A

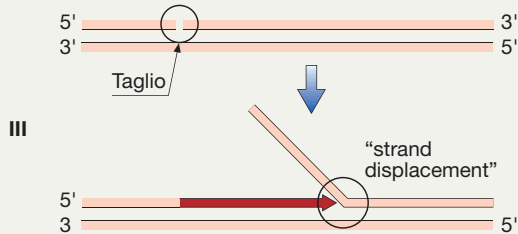
Le DNA polimerasi non sono in grado di iniziare la sintesi di DNA su DNA a doppio o singolo filamento senza un innesco



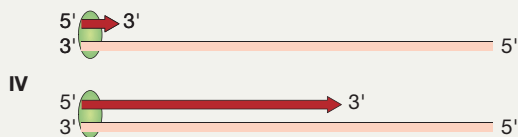
Il 3'-OH è fornito da un innesco di RNA sintetizzato dalla primasi



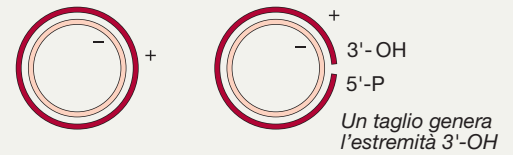
Il DNA a doppio filamento è tagliato per fornire una estremità 3'-OH alla DNA polimerasi



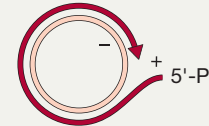
Un 3'-OH di innesco è fornito da una proteina che si lega al DNA

**B**

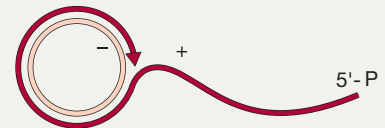
Lo stampo è un DNA circolare a doppio filamento (RF del DNA a singolo filamento di alcuni batteriofagi)



L'allungamento dell'estremità 3'-OH sposta il filamento + formando una specie di coda



Dopo una rivoluzione il filamento spostato raggiunge la lunghezza di una unità



L'allungamento genera un filamento + che contiene molte unità che possono poi venire tagliate per generare singole unità fagiche

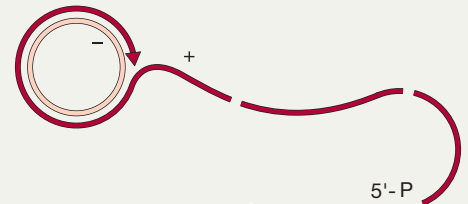
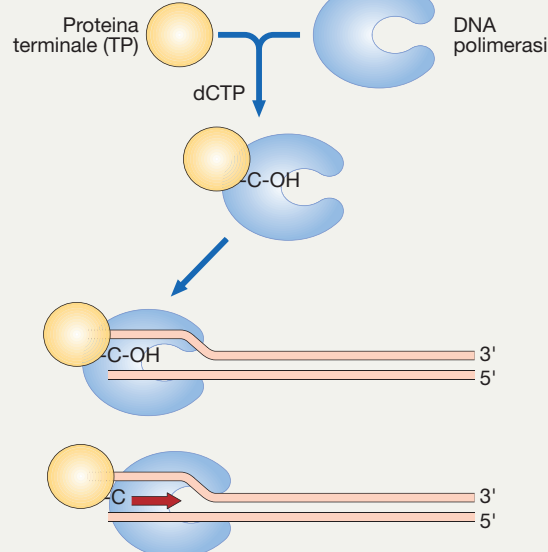
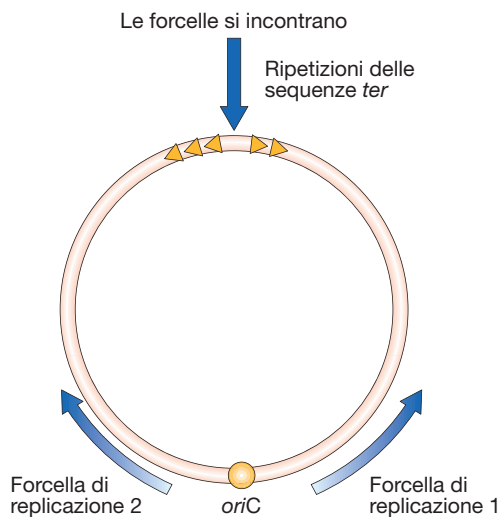
**C**

Figura 6.20 Terminazione della replicazione in *E. coli*. Le due forcelle di replicazione si incontrano a circa 180° rispetto a *oriC*. Le tre ripetizioni di 23 pb della sequenza *ter* a sinistra del punto d'incontro fanno terminare la forcella di replicazione 1, mentre le due ripetizioni *ter* a destra del punto d'incontro fanno terminare la forcella 2. La separazione delle due molecole figlie circolari richiede anche l'azione di DNA topoisomerasi.



6.5 Meccanismo della replicazione del DNA negli eucarioti

La replicazione dei numerosi repliconi presenti sulle grandi molecole di DNA lineari degli eucarioti rispecchia, nelle sue linee generali, il meccanismo che abbiamo appena descritto per la replicazione del genoma di *E. coli*. La macchina proteica è, però, ancora più complessa perché più numerose sono le proteine che costituiscono il macchinario replicativo. Inoltre, il DNA degli eucarioti è impacchettato in una struttura cromatinica complessa (vedi cap. 4) che influenza profondamente il processo replicativo.

Isolamento di proteine replicative negli eucarioti

Come sono state individuate le proteine coinvolte nella replicazione del DNA negli eucarioti? Come descritto per la replicazione di *E. coli*, l'approccio biochimico diretto ha portato all'individuazione e purificazione di alcuni enzimi che devono necessariamente essere coinvolti nella replicazione del DNA (DNA polimerasi, DNA ligasi, DNA topoisomerasi ecc.). Anche negli eucarioti la replicazione procede bidirezionalmente ed è continua su un filamento (leading) e discontinua sull'altro (lagging). La differenza principale è che i frammenti di Okazaki sono più piccoli negli eucarioti (100-200 nucleotidi) e, come già indicato, il meccanismo replicativo procede più lentamente rispetto ai procarioti.

La maggior parte delle proteine coinvolte nella replicazione del DNA eucariotico è stata individuata originariamente nel lievito *Saccharomyces cerevisiae* e in cellule umane. Il lievito è stato soprattutto utile perché essendo un eucariote unicellulare, a cui sono applicabili le stesse strategie genetiche applicate nei batteri, ha permesso l'identificazione di mutanti temperatura-sensibili difettivi nel processo replicativo, utilizzando la stessa strategia descritta sopra per *E. coli*. Ciò ha consentito di stabilire che numerosi geni codificano per proteine richieste per la replicazione del DNA di lievito e per la sua regolazione durante il ciclo cellulare. Proteine replicative sono state purificate con successo da estratti proteici preparati da cellule umane utilizzando strategie biochimiche molto laboriose.

Sistemi di ricostituzione *in vitro* del processo di replicazione

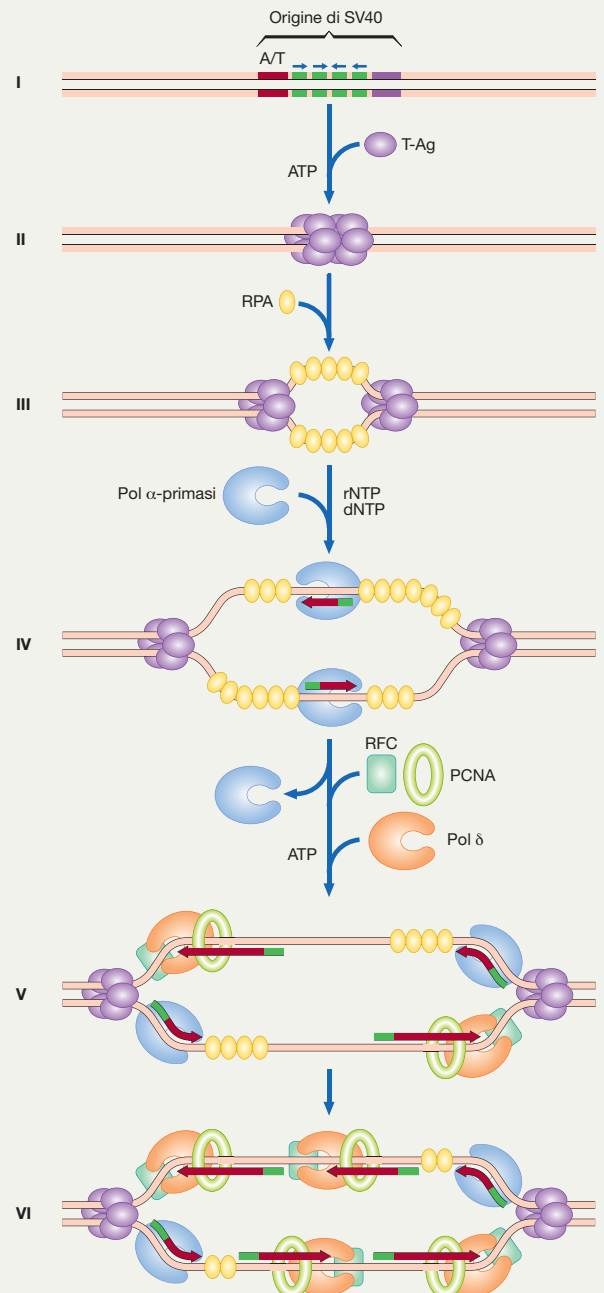
Questo approccio si è principalmente basato sull'utilizzo del DNA di alcuni virus come substrato della reazione. Per esempio, il Simian Virus 40 (SV40) è

un virus che infetta cellule di primati. Il DNA del virus è una piccola molecola di DNA circolare a doppio filamento di circa 5000 pb, che inizia la replicazione da una specifica origine di replicazione di sole 65 pb (**finestra 6.6**). Il DNA del virus codifica per una specifica proteina, chiamata **antigene T** o **T-Ag**, che è assolutamente richiesta per il riconoscimento dell'origine di SV40. Tuttavia, con l'eccezione del T-Ag, all'interno delle cellule infettate dal virus, la replicazione del DNA di SV40 utilizza proteine della cellula ospite.

FINESTRA 6.6

REPLICAZIONE DEL DNA DEL VIRUS SV40 E PROTEINE UMANE RICHIESTE IN TALE PROCESSO

- I. È mostrato uno schema dell'origine di replicazione di SV40 che contiene una regione ricca in AT e corte sequenze ripetute invertite.
- II. Alla sequenza dell'origine virale si legano, in presenza di ATP, due esameri di T-Ag: questa proteina essenziale per la replicazione del DNA virale è richiesta per il riconoscimento dell'origine e ha anche attività di DNA elicasi.
- III. L'azione di T-Ag determina l'apertura della regione dell'origine a cui si lega la proteina *Replication Protein A* (RPA), che è il principale fattore con affinità per ssDNA negli eucarioti.
- IV. L'inizio della sintesi di DNA sulla leading strand richiede l'azione del complesso DNA polimerasi α -primasi (Pol α -primasi) che, in presenza di rNTP e dNTP, porta alla sintesi del primo RNA primer (in verde) e al suo parziale allungamento (in rosso).
- V. A questo stadio avviene una reazione complessa, e ancora in parte non chiarita, che è definita **scambio delle polimerasi**. Infatti, Pol α -primasi viene sostituita dalla DNA polimerasi δ sul 3'OH della catena nascente di DNA. Per effettuare questo scambio intervengono proteine note come PCNA e RFC. PCNA (*Proliferating Cell Nuclear Antigen*) è l'equivalente della β -clamp di *E. coli* e viene caricata, con consumo di ATP, sul DNA tramite l'azione di RFC (*Replication Factor C*), che è l'equivalente del complesso- γ di *E. coli*. Come risultato dello "scambio delle polimerasi", il complesso Pol α -primasi viene spostato vicino alle due forcelle di replicazione di destra e di sinistra dove innesca la sintesi del primo frammento di Okazaki sulla lagging strand di entrambe le forcelle replicative.
- VI. Pol α -primasi aggiunge soltanto 30-40 nucleotidi ai primers di RNA e i frammenti di Okazaki sono portati alla loro dimensione finale (≈ 100 -200 nucleotidi negli eucarioti) dalla DNA polimerasi δ . A ogni scambio di polimerasi sulla lagging strand, Pol α -primasi si sposta in coda iniziando la sintesi del frammento di Okazaki successivo.



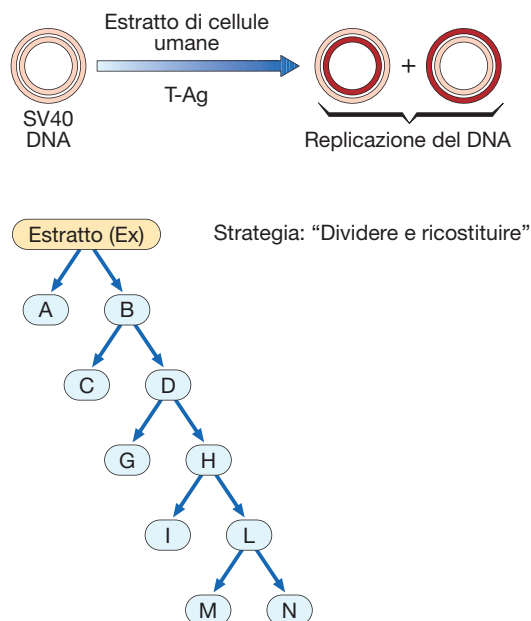


Figura 6.21 Replicazione del DNA del virus SV40 *in vitro*. “Dividere e ricostituire”: la strategia utilizzata per identificare le proteine richieste per replicare il DNA del virus SV40 *in vitro*. Come descritto nel testo, il DNA a doppio filamento circolare del DNA del virus SV40 può essere replicato da un estratto di cellule umane addizionato con la proteina virale nota come antigene-T (T-Ag). Le proteine presenti nell’estratto di cellule umane sono state singolarmente identificate utilizzando lo schema rappresentato nella parte in basso della figura. Brevemente, l’estratto è stato diviso in due parti chiamate A e B. Sia A che B, aggiunte singolarmente al T-Ag, non sono in grado di replicare il DNA di SV-40, suggerendo che le proteine richieste per la replicazione si sono suddivise nelle due frazioni A e B. Si procede, pertanto, a separare B in altre due frazioni, per es. C

e D, mentre si tiene fissa la frazione A. Si provano poi diverse combinazioni: per esempio, solo addizionando A con la frazione D si ottiene replicazione del DNA, indicando che C non contiene proteine d’interesse per il processo. Si continua, quindi, con il dividere D in due frazioni, chiamate G e H, e così via. Si saggiano le diverse combinazioni come indicato sopra: per esempio, soltanto mescolando tra loro le frazioni A, G e I più la proteina purificata N si ottiene la replicazione del virus. A questo punto, tenendo fissa la proteina N, si procede all’isolamento dalle frazioni A, G e I delle proteine presenti in esse e richieste per la replicazione del DNA virale. Alla fine di tutta la procedura (ha richiesto anni di lavoro!) sono state ottenute numerose proteine purificate che, se messe insieme, sono in grado di ricostituire il processo di replicazione di SV40 in provetta.

È stato osservato che, se in una provetta si mette il DNA a doppio filamento del virus e molecole purificate di T-Ag, si riesce a ottenere replicazione del DNA virale aggiungendo un estratto proteico preparato da cellule umane. Quindi, tutti i componenti richiesti per replicare il DNA di SV40, in aggiunta al T-Ag, devono essere presenti nell’estratto proteico di cellule umane.

Come si possono individuare e separare tutte le varie proteine richieste nel processo replicativo? La strategia utilizzata, spesso indicata come “Dividere e ricostituire” è indicata in **figura 6.21**. È una procedura laboriosa che porta di volta in volta alla purificazione di una singola proteina richiesta per la replicazione del DNA di SV40. Ogni volta bisogna ricominciare da capo e procedere alla purificazione di una seconda, poi di una terza proteina e così via. Se il procedimento ha successo, alla fine si ottiene la purificazione di numerose proteine che rimesse insieme in provetta sono in grado di replicare il DNA di SV40 e, quindi, di ricostituire *in vitro* tutto il processo replicativo. Ciò è quello che è stato ottenuto tramite il lavoro decennale di numerosi laboratori: uno schema delle proteine più importanti e dei principali eventi richiesti per la replicazione di SV40 è indicato nella **finestra 6.6**, dove si può osservare che la logica replicativa è pressoché identica a quella descritta per *E. coli*.

La replicazione di SV40 è un modello molto semplificato di quanto avviene negli eucarioti. In **tabella 6.2** sono indicate le più importanti proteine coinvolte nella replicazione dei cromosomi degli eucarioti con l’indicazione della loro struttura e funzione, e il loro confronto con proteine che svolgono un’analoga funzione in *E. coli*. Dalla tabella spiccano alcuni aspetti che meritano di essere sottolineati.

Tabella 6.2 Comparazione tra le proteine più note coinvolte nella replicazione del DNA negli eucarioti e nella replicazione del DNA di *E. coli*.

<i>E. coli</i>	Eucarioti	Struttura e funzione della proteina negli eucarioti
DnaA	ORC (T-Ag in SV40)	Riconoscimento dell'origine; è costituita da 6 subunità
DnaB-DnaC	MCM 2-7 (T-Ag in SV40)	DNA elicasi; è costituita da 6 subunità
DnaG (primasi)	DNA polimerasi α -primasi	La polimerasi che sintetizza gli RNA primer e corte catene di DNA sul filamento ritardato; è composta da 4 subunità
Subunità α di Pol III	DNA polimerasi δ e DNA polimerasi ϵ	Le due polimerasi replicative; ciascuna è composta da 4 subunità
Subunità β di Pol III	PCNA	La proteina che forma la "clamp" che avvolge e scivola sul DNA; è un omotrimerico
Complesso- γ di Pol III	RFC	Il complesso che carica la "clamp" sul DNA; è composto da 5 subunità
SSB	RPA	La proteina che stabilizza il DNA a singolo filamento (ssDNA); è composta da 3 subunità
Pol I	FEN1, RNasi H	Nucleasi che eliminano gli RNA primer dai frammenti di Okazaki
Lig 1	DNA ligasi I	La DNA ligasi che salda i frammenti di Okazaki

Nella tabella non sono indicate le numerose proteine che regolano l'innesco della sintesi del DNA negli eucarioti e collegano il processo replicativo con la progressione del ciclo cellulare (vedi testo). **Abbreviazioni:** ORC: *Origin Recognition Complex*; MCM: *Mini-Chromosome Maintenance*; PCNA: *Proliferating Cell Nuclear Antigen*; RFC: *Replication Factor C*; RPA: *Replication Protein A*; FEN1: *Flap Endonuclease 1*.

1. La maggior parte delle proteine coinvolte nella replicazione sono loro stesse dei complessi proteici che interagiscono ulteriormente tra loro a formare il replisoma eucariotico.
2. Il numero di geni codificanti per proteine replicative è molto più alto negli eucarioti rispetto ai procarioti.
3. La stessa funzione può essere svolta, nei procarioti e negli eucarioti, da proteine solo apparentemente diverse. Per esempio, la subunità β della DNA polimerasi III oloenzima di *E. coli* forma la "clamp" per caricare la polimerasi sul DNA e su questo scorrere durante il processo replicativo. Poco sopra abbiamo detto che la "clamp" di *E. coli* è un omodimero (due subunità uguali tra loro). Se si paragona con metodi bioinformatici la sequenza amminoacidica della subunità β con quella della proteina PCNA degli eucarioti si osserva solo una blanda somiglianza (circa il 20% d'identità tra gli amminoacidi delle due proteine). Quando, però, è stata risolta la struttura cristallografica di PCNA, si è osservato che la proteina è un omotrimerico (tre subunità uguali tra loro) che forma una struttura a ciambella che è praticamente sovrapponibile a quella formata dalla β -clamp. Ciò indica che quello che è stato conservato nel corso dell'evoluzione è la struttura del complesso proteico, anche se la sequenza primaria dei singoli monomeri si è diversificata.
4. Negli eucarioti esistono più DNA polimerasi replicative. Come si osserva in **tabella 6.3**, il numero di DNA polimerasi presenti nelle cellule dei mammiferi è molto più elevato che in *E. coli*. Il fatto che esista, però, una polimerasi coinvolta nell'innesco della sintesi del DNA (il complesso DNA polimerasi α -primasi) e polimerasi (δ e ϵ) coinvolte, rispettivamente, nell'allungamento di questi inneschi e nella sintesi continua della leading strand, implica che durante la replicazione del DNA deve avvenire uno scambio delle polimerasi. Infatti, dopo che la DNA polimerasi α -primasi ha sintetizzato il primer di RNA (8-12 nt) e lo ha allungato di poche decine di nucleotidi (si stacca facilmente dal DNA non interagendo con il PCNA), viene sostituita da pol δ e/o pol ϵ , mentre pol α -primasi, nella sintesi del filamento ritardato, si sposta sul DNA per innescare la sintesi del frammento di Okazaki successivo (**finestra 6.6**). La sostituzione di pol α -primasi con polimerasi più processive (cioè legate più stabilmente al DNA) aumenta la velocità di sintesi e favorisce la fedeltà dell'intero processo: infatti, pol α -primasi, contrariamente a pol δ e pol ϵ , non possiede un'attività esonucleasica 3'→5' con funzione di correzione di bozze.

In **figura 6.22** e in **tabella 6.2** sono indicati i principali ruoli di proteine replicative e le principali differenze tra i meccanismi di replicazione del virus SV40 e ciò che accade sulle forcelle di replicazione attive nei cromosomi degli eucarioti. Le proteine principali sono pressoché identiche: le maggiori differenze riguardano le proteine che riconoscono l'origine e le elicasi e polimerasi coinvolte nel processo.

Sintesi di DNA translesione

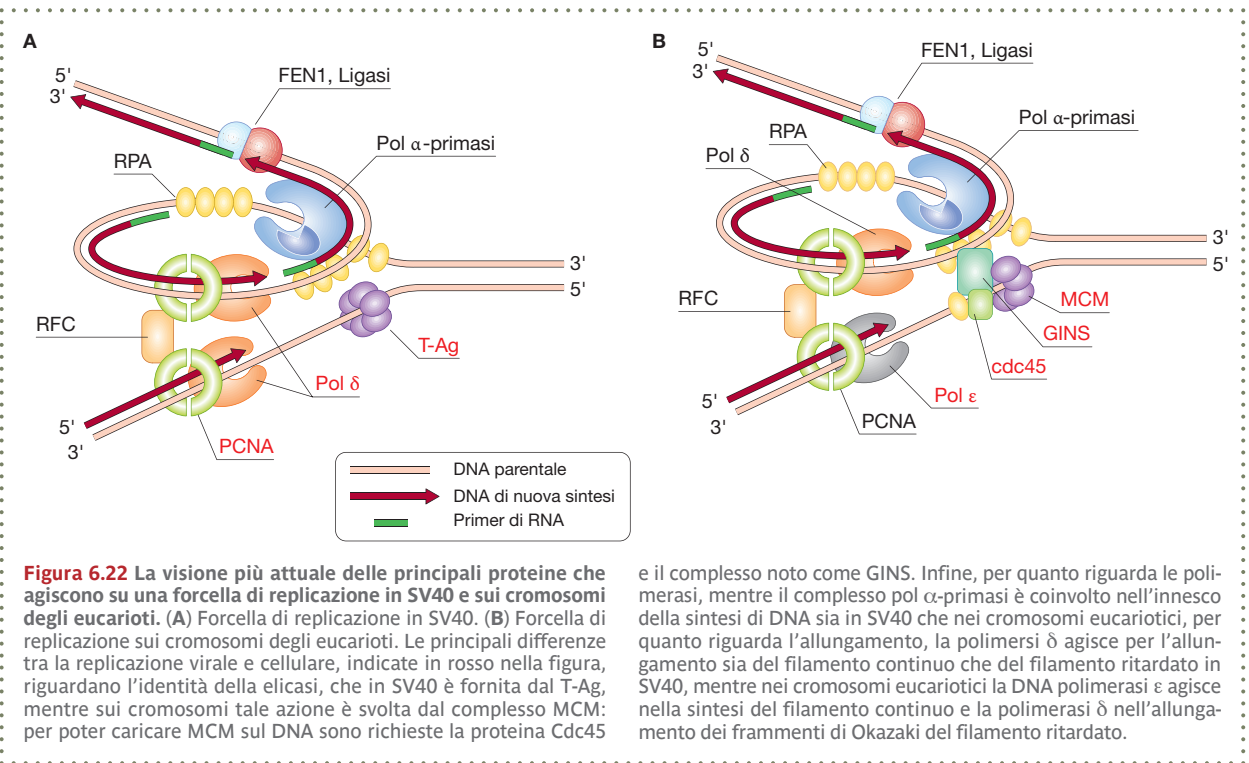
Dall'elenco delle DNA polimerasi presenti nei mammiferi riportato in **tabella 6.3**, si osserva che la maggior parte di esse non è coinvolta nella replicazione vera e propria del DNA, ma nella sua riparazione e nella **sintesi di DNA translesione** (*vedi cap. 7*). Ciò ha una sua logica. Infatti, come vedremo meglio più avanti, il DNA viene alterato da un numero elevatissimo di agenti chimico-fisici che causano danni diversi sul DNA. Queste lesioni sono fortunatamente riparate da una serie di meccanismi, in gran parte specifici per i diversi tipi di danno.

Una lesione sul DNA impedisce la progressione della forcella di replicazione, che si blocca finché il danno non viene riparato. Talvolta, però, la riparazione fallisce e la cellula ha così sviluppato dei meccanismi di tolleranza al danno per cui la sintesi del DNA riesce a procedere oltre la lesione (sintesi di DNA translesione o TLS). Poiché i danni possono essere diversi si sono evolute le numerose TLS DNA polimerasi indicate in **tabella 6.3**. Come vedremo nel capitolo 7, queste polimerasi sostituiscono le polimerasi replicative e sono in grado di posizionare nucleotidi di fronte alle lesioni perché hanno un sito catalitico meno selettivo di quello delle polimerasi replicative. Ne deriva che la sintesi di DNA translesione introduce, nella maggior parte dei casi, degli errori e, quindi, delle mutazioni.

Tabella 6.3 DNA polimerasi eucariotiche.

Nome	Funzione
Pol α (alfa)	Sintesi degli RNA primer e dei primi 20-30 nucleotidi di DNA nei frammenti di Okazaki
Pol δ (delta)	Allungamento dei frammenti di Okazaki Riparazione del DNA (NER, MMR, DSBR)
Pol ϵ (epsilon)	Sintesi del filamento leading Riparazione del DNA (NER, DSBR)
Pol γ (gamma)	Replicazione di DNA mitocondriale
Pol β (beta)	Riparazione del DNA (BER)
Pol η (eta)	Sintesi di DNA translesione (TLS)
Pol ζ (zeta)	Sintesi di DNA translesione (TLS)
Pol θ (teta)	Riparazione degli ICL
Pol ι (iota)	Sintesi di DNA translesione (meiosi)
Pol κ (kappa)	Sintesi di DNA translesione; DSBR
Pol λ (lambda)	Sintesi di DNA translesione
Pol μ (mu)	DSBR
Pol ν (nu)	Riparazione degli ICL (?)
Rev1	Riparazione dei siti abasici
TdT	Transferasi terminale: unica DNA polimerasi che sintetizza DNA senza copiare uno stampo Ricombinazione V(D)J

Abbreviazioni: NER: *Nucleotide Excision Repair* (il sistema di riparazione del DNA che elimina le lesioni causate da radiazioni UV); BER: *Base Excision Repair* (il sistema che ripara numerose lesioni causate da una varietà di agenti chimici alle basi azotate del DNA); MMR: *Mismatch Repair* (il sistema di riparazione del DNA che elimina gli appaiamenti sbagliati tra le basi durante la replicazione); DSBR: *Double Strand Break Repair* (il sistema di riparazione delle rotture a doppio filamento sul DNA); ICL: *Interstrand Cross-Links* (legami covalenti tra i due filamenti di DNA); V(D)J: una regione dei geni codificanti per le immunoglobuline.



e il complesso noto come GINS. Infine, per quanto riguarda le polimerasi, mentre il complesso pol α -primasi è coinvolto nell'innesco della sintesi di DNA sia in SV40 che nei cromosomi eucariotici, per quanto riguarda l'allungamento, la polimerasi δ agisce per l'allungamento sia del filamento continuo che del filamento ritardato in SV40, mentre nei cromosomi eucariotici la DNA polimerasi ϵ agisce nella sintesi del filamento continuo e la polimerasi δ nell'allungamento dei frammenti di Okazaki del filamento ritardato.

6.6 Replicazione dei telomeri nei cromosomi lineari degli eucarioti

L'evoluzione dei cromosomi lineari degli eucarioti ha fatto sorgere un problema nella loro replicazione che non si pone per molecole di DNA circolari come quella di *E. coli*. Il problema è legato al meccanismo della sintesi discontinua del DNA sul filamento ritardato: infatti, come schematizzato in **figura 6.23**, mentre la replicazione del filamento leading continua fino a copiare l'estremità 5', nella replicazione del filamento lagging rimane il tratto dell'RNA primer al 5' che, una volta rimosso, lascia un segmento di DNA non replicato che non può essere riempito da alcuna delle polimerasi conosciute.

Le regioni terminali dei cromosomi eucariotici sono note come **telomeri** e la loro struttura è molto importante per il controllo della stabilità del genoma. Sui telomeri si formano, infatti, dei complessi nucleoproteici molto caratteristici (vedi cap. 4) che li distinguono da rotture della doppia elica (DSB, *Double Strand Break*), a cui la cellula cerca di porre rimedio attivando meccanismi riparativi (vedi cap. 7) che sarebbero molto dannosi se applicati alle estremità telomeriche.

È importante mantenere nella cellula l'integrità dei telomeri e replicarli correttamente senza perdere informazione genetica. È stato dimostrato che se non si provvede a replicare il tratto telomerico corrispondente all'estremità 5' del filamento lagging descritto sopra, i telomeri umani perdono 50-150 pb alle estremità a ogni divisione cellulare, suggerendo, quindi, che esiste un attivo meccanismo di erosione dei telomeri non replicati correttamente. In alcuni organismi (ad es. *Drosophila melanogaster*) il mantenimento dei telomeri avviene mediante meccanismi ricombinativi. Nella maggior parte dei casi, tuttavia, oltre a un'eventuale azione della ricombinazione, il mantenimento dei telomeri è assicurato dall'azione di un particolare enzima, chiamato **telomerasi**.

Prima di descrivere il meccanismo d'azione della telomerasi bisogna ricordare quanto descritto nel capitolo 4. I telomeri sono costituiti da sequenze di DNA

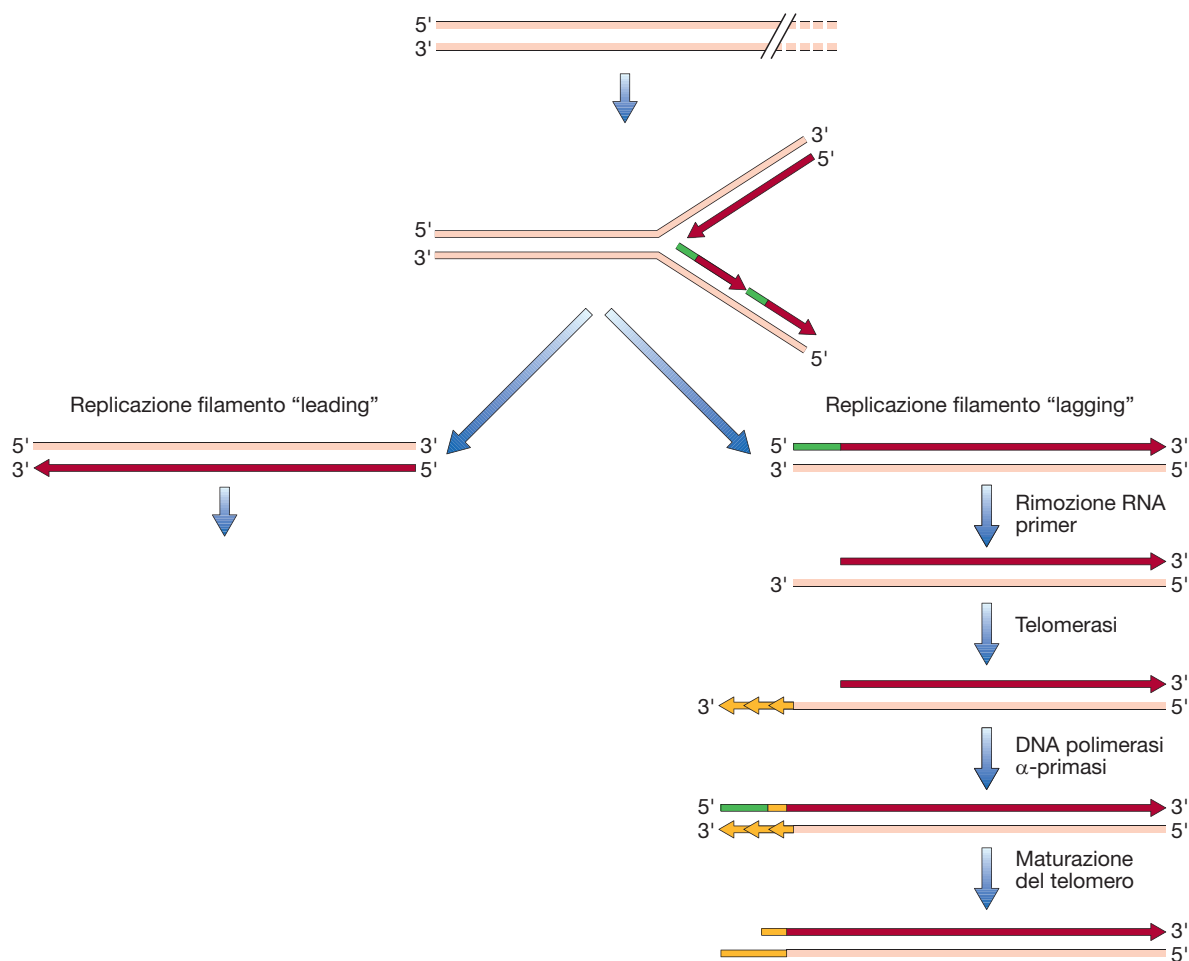


Figura 6.23 Problema della replicazione dei telomeri sui cromosomi lineari. Come si osserva nella figura, la sintesi discontinua del DNA sulla lagging strand genera, dopo rimozione dell'RNA primer, dei buchi non replicati alle estremità telomeriche. Tali buchi non possono essere riempiti da alcuna delle DNA polimerasi cono-

sciute, dato che tutte sono solo in grado di aggiungere nucleotidi solo all'estremità 3'OH. L'azione della telomerasi (vedi fig. 6.24) risolve il problema mediante l'allungamento delle estremità 3' (punte di freccia gialle) e la successiva azione della DNA polimerasi α -primasi su un telomero allungato.

molto particolari: contengono numerose (decine o centinaia) ripetizioni in tandem di brevi sequenze ricche in G che variano leggermente da specie a specie (vedi tab. 4.2) e che lasciano un'estremità sporgente in 3'. Nell'uomo la ripetizione che sporge in 3' ha la sequenza 5'-TTAGGG-3'. La struttura e il meccanismo d'azione della telomerasi sono molto particolari e poco intuitivi.

La telomerasi è, infatti, una ribonucleoproteina costituita da due componenti principali:

1. una proteina chiamata **TERT** (*Telomerase Reverse Transcriptase*) che agisce come una trascrittasi inversa, essendo capace di sintetizzare DNA copiando uno stampo di RNA;
2. una molecola di RNA stampo, chiamata **TERC** (*Telomerase RNA Component*).

La subunità catalitica TERT si associa con altre proteine accessorie (chiamate generalmente Est, "ever shorter telomeres") a formare la macchina proteica coinvolta nel mantenimento dei telomeri (fig. 6.24).

Contrariamente a quanto ci si potrebbe aspettare, la telomerasi non estende il filamento 5' ritardato più corto, ma, al contrario, allunga ulteriormente il terminale 3' sporgente (fig. 6.24). Per fare questo, il terminale 3' sporgente del telome-

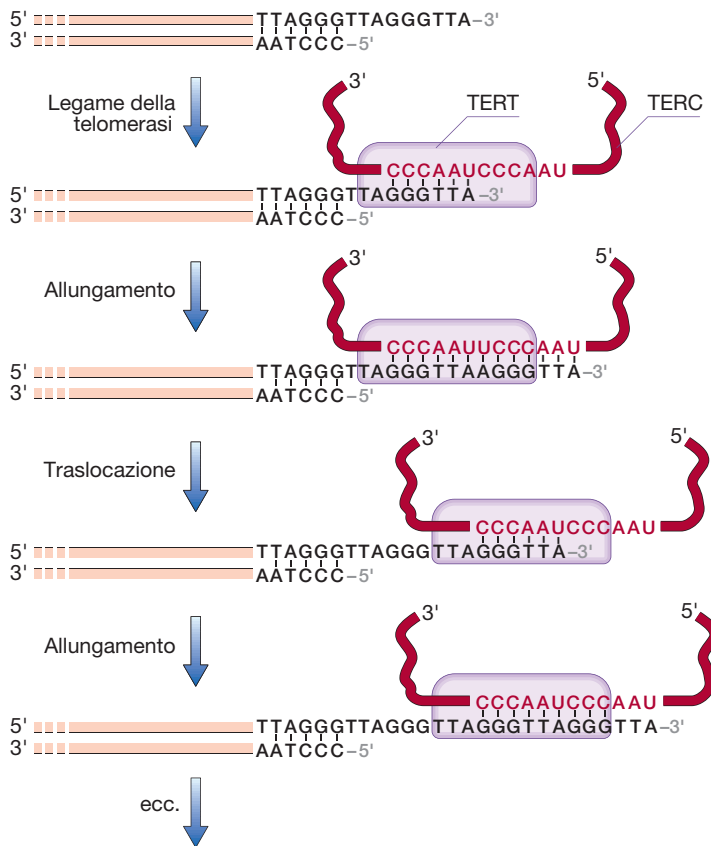


Figura 6.24 Meccanismo d'azione della telomerasi umana. I vari passaggi della reazione di allungamento del telomero da parte della telomerasi sono discussi nel testo.

ro si appaia con il TERC della telomerasi, che ha una sequenza complementare alla sequenza sporgente telomerica: la parte proteica della telomerasi copia poi la sequenza del suo RNA allungando, quindi, di una ripetizione la sequenza 3' sporgente. La successiva traslocazione e il riposizionamento della telomerasi fanno sì che questo processo possa ripetersi più volte determinando l'aggiunta di numerose unità della sequenza telomerica ripetuta. Per procedere alla ricostituzione del telomero interviene ora la DNA polimerasi α -primasi, che, avendo a disposizione uno stampo più lungo, copia il filamento sporgente in 3' esteso dalla telomerasi. È chiaro che il problema originale si ripropone perché il primer di RNA utilizzato da pol α -primasi viene nuovamente rimosso. L'estensione delle sequenze ripetute da parte della telomerasi permette, però, di raggiungere un'omeostasi della lunghezza dei telomeri che non ne altera la funzionalità e impedisce l'erosione d'informazione genetica a partire dall'estremità dei cromosomi. La stabilità dei telomeri è anche influenzata dalla sua struttura. Infatti, come abbiamo discusso nel capitolo 4 (*vedi* par. 4.7), l'estremità 3' sporgente può appaiarsi con la sequenza complementare delle unità ripetute creando una regione ad ansa del DNA osservabile al microscopio elettronico (*vedi* fig. 4.12). Tale struttura può influenzare sia la stabilità del telomero che l'accesso della telomerasi all'estremità 3'.

In molti organismi, come per esempio in lievito, la telomerasi è sempre espressa, mentre nell'uomo la sua espressione ad alti livelli è limitata a cellule che proliferano attivamente. È stata osservata, inoltre, una possibile correlazione tra invecchiamento cellulare e accorciamento dei telomeri, così come il fatto che cellule tumorali riattivano l'espressione della telomerasi. Quindi, da una parte, la telomerasi può influenzare l'invecchiamento e, dall'altra, è attivamente studiata per possibili applicazioni antitumorali.

Replicazione del DNA nella sua struttura cromatinica

La replicazione del DNA, così come anche tutti gli altri aspetti del metabolismo del DNA, sono negli eucarioti fortemente influenzati dalla struttura della cromatina (*vedi* cap. 4). Una forcella di replicazione che avanza sul DNA deve passare attraverso la struttura cromatinica del DNA parentale tramite un meccanismo che non è ancora ben chiarito; è però noto che fattori di rimodellamento della cromatina, che verranno descritti nel capitolo 17 (*vedi* par. 17.2), sono necessari per replicare con efficienza il DNA.

Un aspetto importante da considerare riguarda il destino dei **nucleosomi**, cioè dell'ottamero di istoni, presenti sul DNA parentale che viene duplicato. La maggior parte degli istoni che erano presenti sul DNA parentale viene distribuita sulle due eliche figlie; tuttavia, il DNA è duplicato ed è quindi necessario anche un raddoppio delle molecole istoniche per riformare la struttura nucleosomica sulle molecole figlie. Questa notevole quantità di istoni, approssimativamente uguale in massa al nuovo DNA sintetizzato, è assicurata dal fatto che i geni codificanti per gli istoni sono ripetuti circa 20 volte nel genoma dei vertebrati. Tali geni sono specificamente trascritti nella fase S e il loro mRNA è rapidamente degradato alla fine della stessa fase, mentre le proteine istoniche sono molto stabili. Via via che le forcelle di replicazione procedono sulla cromatina, l'ottamero di istoni viene disassemblato in un tetramero H3-H4 e due dimeri H2A-H2B (**fig. 6.25**). Il tetramero H3-H4 resta associato al DNA e si distribuisce casualmente sulle due doppie eliche figlie, mentre i dimeri H2A-H2B si dissociano dal DNA. Tetrameri H3-H4 di neosintesi si assemblano sulle due doppie eliche figlie negli spazi lasciati vuoti, e i dimeri H2A-H2B (metà vecchi e metà neosintetizzati) sono aggiunti casualmente per riformare la struttura nucleosomica completa. L'assemblaggio dei nucleosomi sulle molecole figlie avviene non appena c'è abbastanza spazio sul DNA per ricostituire un nucleosoma completo (circa 200 pb) e necessita di fattori proteici che favoriscono l'assemblaggio, tra cui i più noti sono chiamati CAF1 e NAP1.

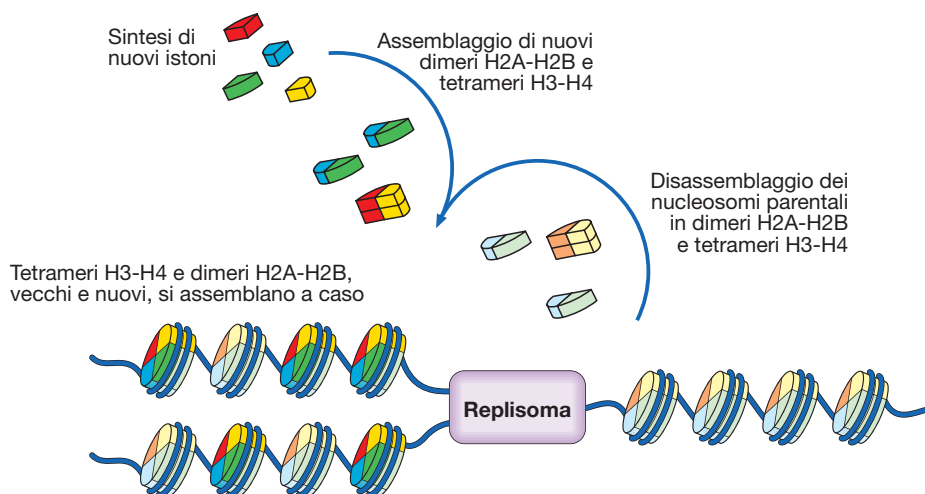


Figura 6.25 Conservazione distributiva dei nucleosomi su una forcella di replicazione. Durante l'avanzamento della forcella di replicazione i nucleosomi sul DNA non replicato si dissociano e si riasssemblano sulle due molecole figlie di DNA. I tetrameri degli istoni H3 e H4 presenti sul DNA parentale non si dissociano e si distribuiscono casualmente sulle due molecole figlie analogamente a quanto succede per i tetrameri H3-H4 di neosintesi. I tetra-

meri H2A-H2B parentali, invece, si dissociano in dimeri: i dimeri vecchi possono combinarsi tra di loro o con dimeri di neosintesi, e i tetrameri così generati si combinano casualmente con i tetrameri H3-H4 vecchi o nuovi. Ne deriva che le molecole figlie di DNA posseggono una miscela casuale di tetrameri H3-H4 e dimeri H2A-H2B di nuova e vecchia sintesi.

Come discusso più in dettaglio nel capitolo 19 (*vedi* par. 19.6), le modificazioni istoniche sugli istoni H3 e H4 ereditati casualmente sulle due eliche figlie di DNA sono uno dei meccanismi alla base dell'ereditarietà epigenetica.

6.7 Controllo della replicazione del DNA durante il ciclo cellulare

La replicazione del DNA è limitata a una fase specifica del ciclo cellulare eucariotico, chiamata fase S. Il processo replicativo è quindi regolato e influenzato dai meccanismi che controllano la progressione del ciclo. Per affrontare il problema della regolazione della replicazione del DNA negli eucarioti è quindi necessario descrivere almeno brevemente il ciclo cellulare.

Aspetti generali del ciclo cellulare

Per dividersi e proliferare, tutte le cellule eucariotiche devono eseguire correttamente un programma genetico, definito **ciclo cellulare**, che sovrintende alla corretta replicazione e segregazione del materiale ereditario nelle cellule figlie. I cromosomi di una cellula eucariotica vengono duplicati in un ristretto intervallo temporale del ciclo cellulare, definito **fase S**, e segregano nelle cellule figlie in mitosi (**fase M**). La fase S e la fase M sono separate da due altre fasi, chiamate **G1** e **G2** (**fig. 6.26A**). È importante che ciascun segmento del genoma sia replicato una volta e una volta sola durante la fase S; esistono dei meccanismi che controllano che la replicazione sia avvenuta correttamente e sia terminata prima che i cromosomi duplicati possano iniziare a separarsi durante la mitosi. Tali meccanismi sono chiamati “checkpoint” e sono discussi poco più avanti.

La duplicazione dei cromosomi e la loro separazione nelle cellule figlie sono processi periodici estremamente precisi e ogni alterazione in questo processo provoca conseguenze catastrofiche in quanto, alterando il numero o l'integrità dei cromosomi, si modifica l'informazione genetica della cellula. Al contrario, la maggior parte dei componenti cellulari extracromosomici (proteine, ribosomi, organelli ecc.) sono presenti in numero elevato e variabile all'interno della cellula e la loro sintesi, correlata alla crescita delle dimensioni della cellula, è pressoché continua durante tutto il ciclo cellulare. Ne deriva che i meccanismi di controllo del ciclo devono essere in grado di integrare tra di loro i numerosi processi biochimici coinvolti sia nella crescita cellulare che nella corretta duplicazione e separazione dei cromosomi.

L'utilizzo di organismi modello (*vedi* cap. 22), specialmente i lieviti *Saccharomyces cerevisiae* e *Schizosaccharomyces pombe*, è stato essenziale per individuare i geni che codificano per proteine richieste per la corretta progressione del ciclo cellulare: tali geni sono indicati come **geni CDC** (**Cell Division Cycle**). Poiché l'incapacità di progredire attraverso il ciclo cellulare è letale, Leland Hartwell e Paul Nurse (i pionieri di tali ricerche) hanno cercato mutazioni che bloccassero la progressione del ciclo cellulare solo in particolari condizioni (**mutazioni condizionali**), così da poterne studiare le conseguenze trasferendo i mutanti da condizioni “permissive” (in genere 23 °C in lievito) a condizioni “non permissive” (in genere 37 °C); abbiamo già descritto in questo capitolo l'utilità di tali mutazioni. L'individuazione di mutanti *cdc* in *S. cerevisiae* è stata facilitata dal fatto che questo lievito possiede una serie di parametri morfologici che possono essere facilmente seguiti durante il ciclo cellulare (**fig. 6.26B**). L'inizio della formazione della gemma (che poi diventerà la cellula figlia) coincide, in prima approssimazione, con l'inizio della fase S ed è preceduto dalla duplicazione del centro di organizzazione del fuso (*Spindle Pole Body*, SPB) e seguito dalla migrazione nucleare, dall'allungamento delle fibre del fuso mitotico, dalla divisione nucleare e dalla citocinesi. Mutazioni condizionali nei geni *CDC* di *S. cerevisiae* sono identifica-

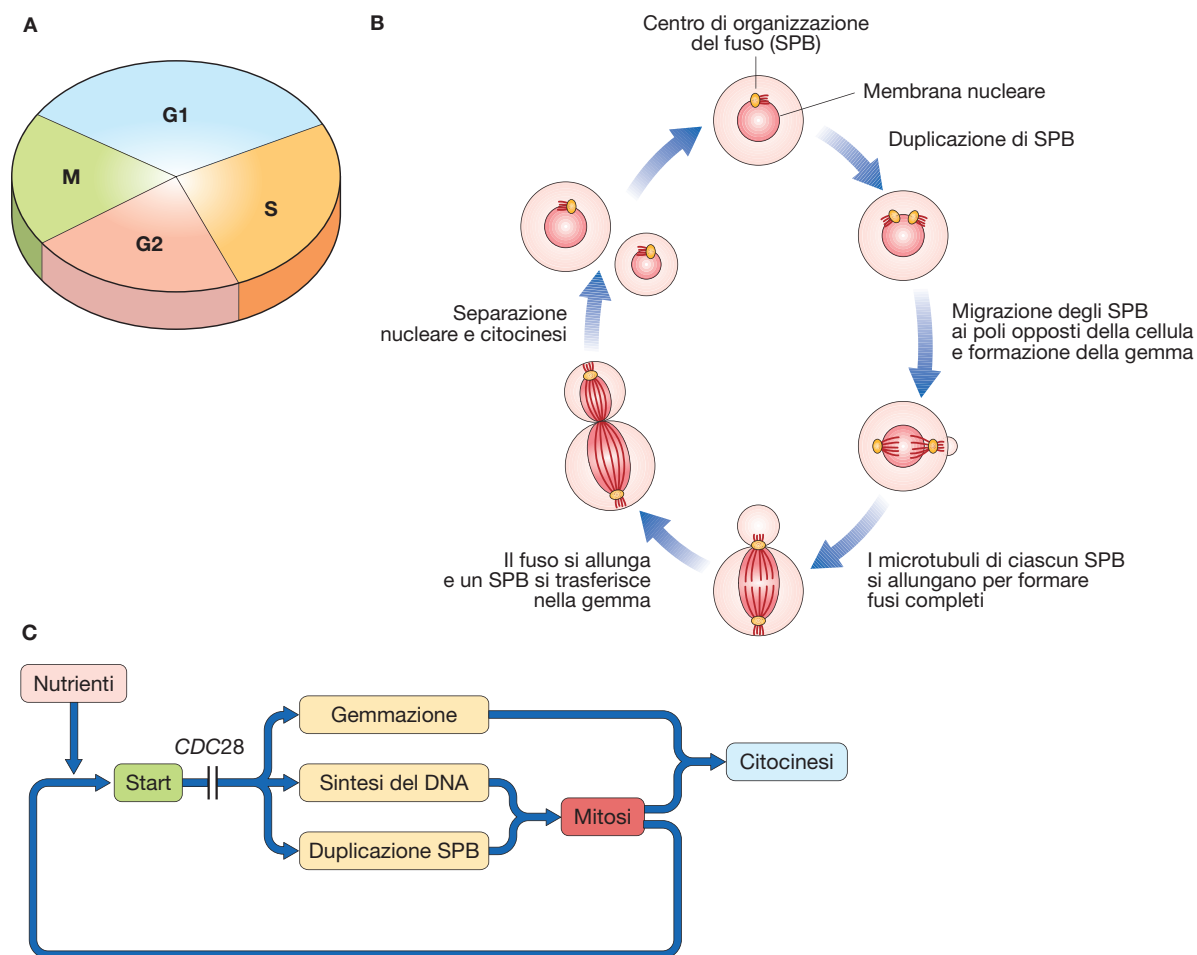


Figura 6.26 Ciclo cellulare. (A) Schema generale del ciclo cellulare. (B) In *S. cerevisiae* le varie fasi del ciclo cellulare possono venire distinte seguendo alcuni parametri morfologici indicati in figura. (C) Il raggiungimento di una certa massa critica alla fine della fase G1 (assicurata dalla presenza di appropriati nutrienti nel terreno di crescita) permette alle cellule di lievito di superare un punto di

controllo molto importante, chiamato "Start" (nelle cellule umane esiste un punto di controllo analogo, chiamato "Restriction Point"). Il superamento di Start è controllato, in lievito, dal gene *CDC28* e innesca una serie di eventi (indicati in figura) che portano infine alla mitosi e alla divisione cellulare (citocinesi).

bili abbastanza facilmente perché le cellule mutanti si arrestano tutte con la stessa morfologia (fenotipo terminale), dopo trasferimento alle condizioni non permissive, in quanto non può svolgersi una funzione richiesta per il superamento di uno specifico passaggio del ciclo cellulare. La caratterizzazione dei mutanti *cdc* di lievito ha permesso di stabilire un aspetto estremamente importante nel controllo del ciclo cellulare: l'esecuzione di eventi tardivi dipende dalla corretta esecuzione degli eventi precedenti. Per esempio, una mutazione che blocca la sintesi del DNA determina l'arresto del ciclo cellulare nella fase S e previene l'entrata delle cellule in mitosi e la successiva divisione cellulare. Ne deriva che la mitosi e la divisione cellulare richiedono il precedente completamento della fase S e sono quindi organizzate come una serie lineare di eventi tra loro interdipendenti. La visione del ciclo cellulare come una cascata di eventi tra loro interconnessi è risultata applicabile non solo al lievito, ma anche agli altri eucarioti.

Il gene di *S. cerevisiae* denominato *CDC28* svolge un ruolo particolarmente rilevante nel controllo del ciclo cellulare. Infatti, la maggior parte delle mutazioni in questo gene altera l'esecuzione, verso la fine della fase G1, di un evento centrale del ciclo cellulare di lievito, chiamato **Start** (fig. 6.26C), superato il quale la

cellula è irreversibilmente programmata a entrare nel ciclo mitotico. Sorprendentemente, fu trovato che la funzione del gene *CDC28* può essere svolta dal gene *CDC2* umano se questo viene espresso in cellule di lievito. Questo dato eclatante dà un'idea di quanto importanti e conservate nel corso dell'evoluzione possano essere le funzioni di geni richiesti per la progressione del ciclo cellulare.

Senza entrare troppo nei dettagli, i processi molecolari che permettono la corretta esecuzione del ciclo cellulare possono essere così raggruppati, come indicato in **figura 6.27A**:

1. esiste “un motore” del ciclo cellulare, costituito da enzimi di cui il prodotto del gene *CDC28* è il prototipo;
2. lo stato e il funzionamento del motore variano durante il ciclo cellulare, determinando una cascata di eventi “a valle” che sono richiesti per eseguire le diverse transizioni del ciclo cellulare (per esempio il passaggio dalla fase G1 alla fase S, il passaggio dalla fase G2 alla fase M, o l'uscita dalla mitosi e il conseguente ritorno in G1);
3. lo stato e il funzionamento del motore vengono controllati da meccanismi di sorveglianza, chiamati **checkpoint**, che valutano se sia realmente opportuno passare da una fase all'altra del ciclo cellulare.

Questi meccanismi di sorveglianza possono essere “intrinseci” o “estrinseci” al ciclo cellulare stesso, nel senso che controllano il passaggio da una fase all'altra del ciclo sia in condizioni fisiologiche che patologiche. Per esempio, una cellula non può entrare nella fase S se non ha superato un punto critico (che corrisponde a Start in cellule di lievito ed è chiamato **Restriction Point** o **R** nelle cellule di mammifero), e il superamento di tale punto dipende dal raggiungimento di una massa cellulare critica. Un meccanismo analogo controlla l'entrata in mitosi. Questi meccanismi di sorveglianza intrinseci assicurano la corretta successione dei vari passaggi del ciclo cellulare in condizioni fisiologiche. I meccanismi di sorveglianza estrinseci, invece, modulano la risposta della cellula in condizioni “patologiche”, per esempio quando il DNA viene danneggiato: di tale checkpoint e delle conseguenze di un suo mal funzionamento discuteremo più avanti nel capitolo 7 (*vedi* par. 7.3).

Motore del ciclo cellulare

Il motore del ciclo cellulare è costituito da una famiglia di proteine chinasi, enzimi in grado di trasferire un gruppo fosfato dall'ATP a specifici amminoacidi di altre proteine. Tali chinasi sono essenzialmente prive di attività enzimatica nella loro forma monomerica e, per essere funzionali, devono associarsi ad altre proteine, chiamate **cicline**. Il loro nome deriva dal fatto che le cicline sono sintetizzate e degradate in modo ciclico all'interno della cellula e, di conseguenza, il loro livello varia considerevolmente nelle diverse fasi del ciclo cellulare. È proprio in virtù dell'associazione richiesta con le cicline che si parla di **chinasi ciclina-dipendenti**, o **CDK** (*Cyclin-Dependent Kinase*).

In *S. cerevisiae* esiste una sola CDK (il prodotto del gene *CDC28*), mentre nei mammiferi esistono numerose CDK. Sia in lievito che negli eucarioti multicellulari esistono poi numerose cicline, i cui singoli livelli quantitativi variano a stadi diversi del ciclo cellulare come risultato sia di una sintesi che di una degradazione differenziale. Alcune cicline raggiungono, quindi, il loro massimo livello durante la fase G1, altre durante la fase S e altre ancora durante la fase G2. Poiché l'attività delle CDK richiede la preventiva associazione con l'una o l'altra delle diverse cicline e i substrati che vengono fosforilati variano a seconda del tipo di ciclina associata alle CDK, ne deriva che sia il livello quantitativo che il tipo di ciclina che viene prodotto a stadi diversi del ciclo cellulare sono dei fattori essenziali che determinano il passaggio da una fase all'altra del ciclo. Il motore del ciclo cellulare è, quindi, essenzialmente lo stesso in cellule di lievito e di mammifero: l'unica

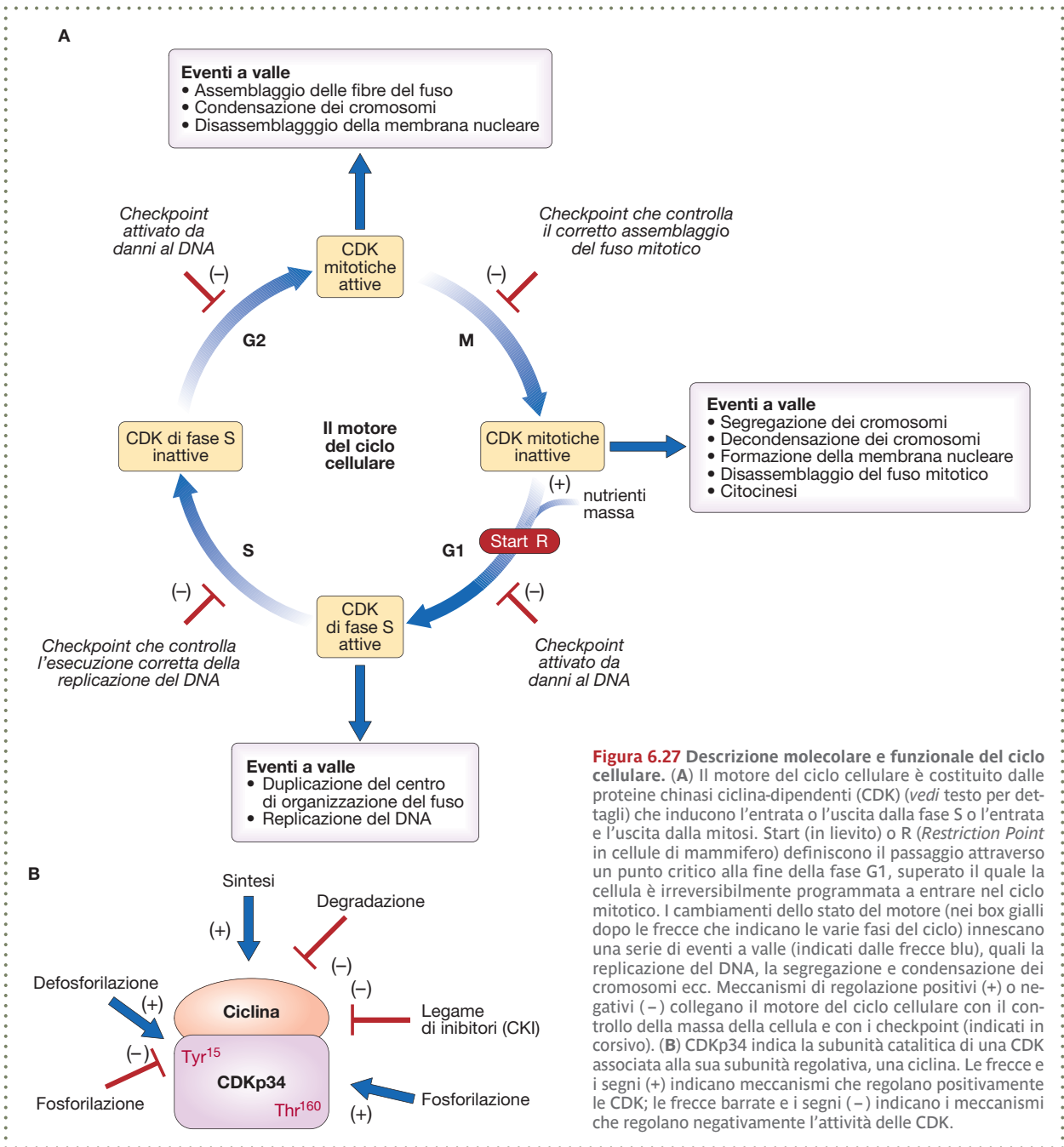


Figura 6.27 Descrizione molecolare e funzionale del ciclo cellulare. (A) Il motore del ciclo cellulare è costituito dalle proteine chinasi ciclina-dipendenti (CDK) (vedi testo per dettagli) che inducono l'entrata o l'uscita dalla fase S o l'entrata e l'uscita dalla mitosi. Start (in lievito) o R (Restriction Point in cellule di mammifero) definiscono il passaggio attraverso un punto critico alla fine della fase G1, superato il quale la cellula è irreversibilmente programmata a entrare nel ciclo mitotico. I cambiamenti dello stato del motore (nei box gialli) dopo le frecce che indicano le varie fasi del ciclo) innescano una serie di eventi a valle (indicatedi dalle frecce blu), quali la replicazione del DNA, la segregazione e condensazione dei cromosomi ecc. Meccanismi di regolazione positivi (+) o negativi (-) collegano il motore del ciclo cellulare con il controllo della massa della cellula e con i checkpoint (indicatedi in corsivo). (B) CDKp34 indica la subunità catalitica di una CDK associata alla sua subunità regolativa, una ciclina. Le frecce e i segni (+) indicano meccanismi che regolano positivamente le CDK; le frecce barrate e i segni (-) indicano i meccanismi che regolano negativamente l'attività delle CDK.

differenza è che, come abbiamo detto sopra, in lievito c'è una sola CDK mentre negli eucarioti multicellulari ce ne sono varie. Sia in un caso che nell'altro, tuttavia, la diversa specificità d'azione delle CDK è assicurata dall'associazione con diverse cicline, che sono altrettanto numerose in eucarioti uni- e multicellulari.

Come schematizzato in **figura 6.27B**, il legame con le cicline non è l'unico fattore che influenza l'attività delle CDK. Esistono almeno altri due livelli di regolazione altrettanto importanti:

1. la presenza di inibitori;
2. lo stato di fosforilazione stesso delle CDK.

Sia in lievito che in organismi multicellulari sono state scoperte una serie di proteine che agiscono come **inibitori delle CDK** che vengono chiamate **CKI** (*Cyclin-De-*

pendent Kinase Inhibitors). Se le CDK possono essere considerate come il “motore” della “macchina” del ciclo cellulare, le proteine CKI possono esserne considerate i “freni”. Alcune CKI sembrano controllare passaggi intrinseci del ciclo cellulare, mentre altre sembrano bloccare la proliferazione cellulare in risposta a stimoli esterni. Va infatti ricordato che il ciclo cellulare è anche influenzato da segnali esterni, quali fattori di crescita e antimitogeni, al fine di integrare la divisione cellulare sia con l'ambiente extracellulare che con i meccanismi alla base del differenziamento. Alcune CKI sembrano avere un ruolo essenziale proprio nello stabilire queste interconnessioni. Le stesse CKI, o analoghe proteine con funzione inibitoria, possono essere ugualmente importanti per completare i programmi dello sviluppo embrionale o per bloccare la crescita una volta che il programma sia stato completato. È opportuno ricordare che la maggior parte delle cellule dei tessuti di un organismo multicellulare non sono e non debbono essere in fase di attiva proliferazione. È quindi necessario che le cellule non proliferanti di questi tessuti escano permanentemente dal ciclo cellulare ed entrino in una fase di quiescenza (definita come **G0**). Un meccanismo che determini e mantenga un'elevata espressione degli inibitori CKI potrebbe essere idoneo a tale scopo.

Il terzo livello di regolazione dello stato del motore del ciclo cellulare coinvolge il livello di fosforilazione delle CDK (**fig. 6.27B**). Le CDK sono delle proteine chinasi che fosforilano altre proteine, ma la loro attività è modulata dallo stato di fosforilazione della stessa subunità catalitica delle CDK (nota anche come p34). All'interno della proteina p34 esistono dei residui amminoacidici che devono essere fosforilati e altri che devono essere de-fosforilati affinché la proteina p34 sia enzimaticamente attiva, e questi residui sono conservati in tutte le CDK degli eucarioti. In particolare, il residuo di tirosina in posizione 15 (Tyr15) deve essere de-fosforilato, mentre il residuo di treonina in posizione 160 (Thr160) deve essere fosforilato. Da queste considerazioni appare evidente che altre proteine chinasi e proteine fosfatasi (proteine che de-fosforilando specifici amminoacidi svolgono una funzione contraria rispetto alle chinasi) regolano positivamente o negativamente l'attività delle CDK. Come schematizzato in **figura 6.27B**, la proteina chinasi che fosforila il residuo di Tyr15 agisce come regolatore negativo inibendo l'attività delle CDK, mentre la proteina fosfatasi e la proteina chinasi che, rispettivamente, de-fosforilano il residuo Tyr15 e fosforilano il residuo Thr160, possono essere considerate come regolatori positivi.

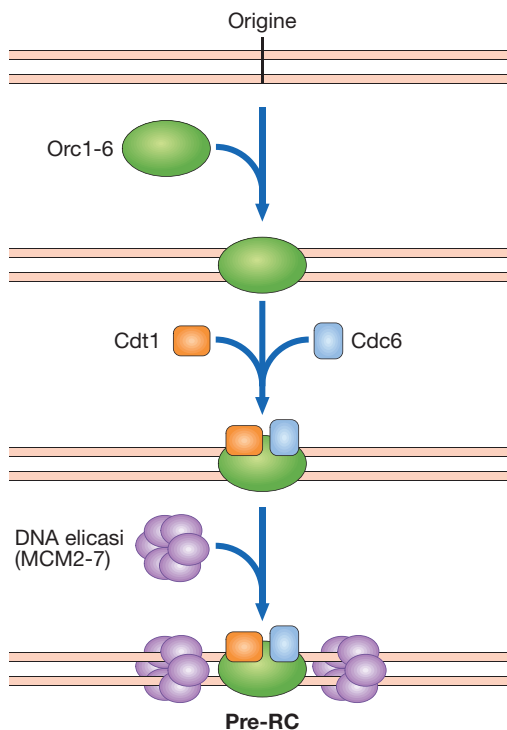
Meccanismi molecolari che assicurano che la replicazione del DNA possa avvenire una volta sola ogni ciclo cellulare

Dopo questa breve descrizione generale del ciclo cellulare e del ruolo delle CDK, vediamo come l'attività di queste proteine chinasi controlla l'accensione delle origini di replicazione assicurando che ogni regione del genoma sia replicata una volta soltanto durante ogni ciclo cellulare.

La selezione e preparazione delle origini di replicazione del DNA negli eucarioti inizia quando le cellule escono dalla mitosi e prosegue nella fase G1 del ciclo cellulare. In questo lasso di tempo si forma su ciascuna origine un complesso proteico definitivo **pre-RC** (*pre-Replication Complex*). Il costituente principale del pre-RC è il complesso ORC (**tab. 6.2**), che, in lievito, sembra stia sempre attaccato alle ARS durante tutto il ciclo cellulare, costituendo quindi una zona libera da nucleosomi, mentre, in altri organismi, si assembla sulle origini quando la struttura del cromosoma si decondensa uscendo dalla mitosi. Il pre-RC comincia a formarsi quando al complesso ORC si agganciano due proteine chiave nel controllo replicativo, chiamate **Cdc6** e **Cdt1**; queste sono richieste per il caricamento del complesso dell'elicasi, nota come MCM2-7 (**fig. 6.28**).

L'attività delle proteine chinasi CDK (*Cyclin-Dependent Kinase*) (vedi sopra) trasforma il pre-RC in quello che è chiamato **pre-initiation complex**, cioè una

Figura 6.28 Fornazione del complesso di pre-inizio (pre-RC o *pre-Replication Complex*). Il complesso ORC si lega alle origini di replicazione e recluta le proteine Cdt1 e Cdc6, che, a loro volta, sono richieste per reclutare il complesso elicastico formato dalle proteine MCM.



Pre-RC

Attivazione della CDK

MCM10

Cdc45

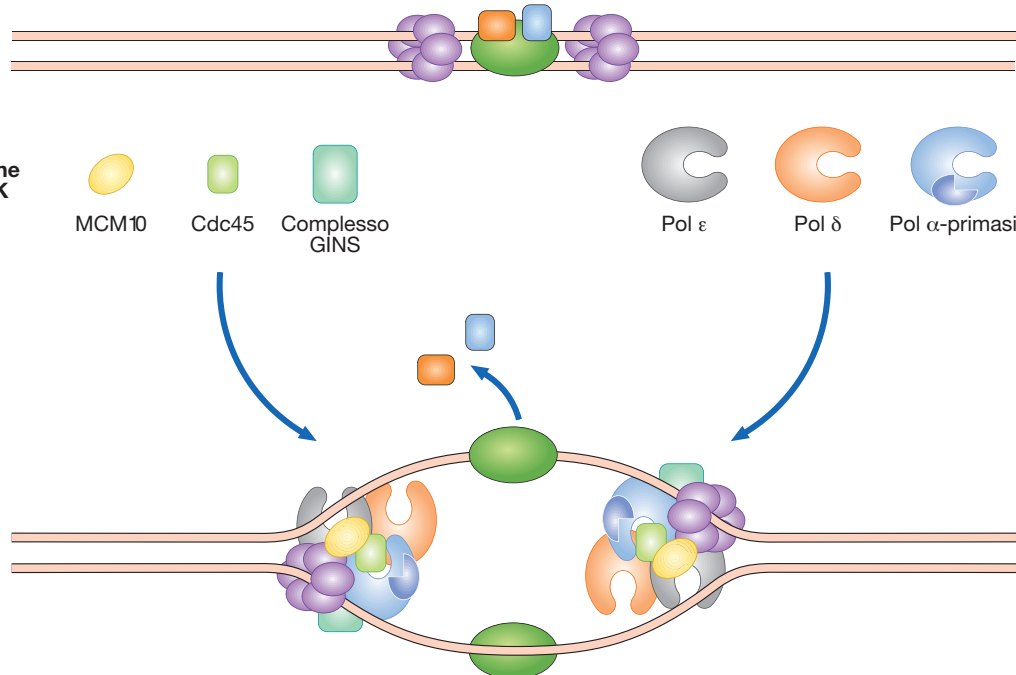
Complesso GINS

Pol ϵ

Pol δ

Pol α -primasi

Post-RC



L'attivazione dell'elicasi MCM innesca tutte le reazioni che fanno partire la sintesi di DNA

Figura 6.29 Passaggio dal pre-RC al post-RC. Tale passaggio richiede principalmente l'attività chinastica delle CDK che porta, da una parte, al distacco delle proteine Cdt1 e Cdc6 dalle origini e, dall'altra, all'attacco di numerose altre proteine che danno origi-

ne al *Post-Replication Complex*. L'attivazione dell'attività elicastica del complesso MCM innesca tutte le reazioni biochimiche che danno inizio alla sintesi vera e propria del DNA.

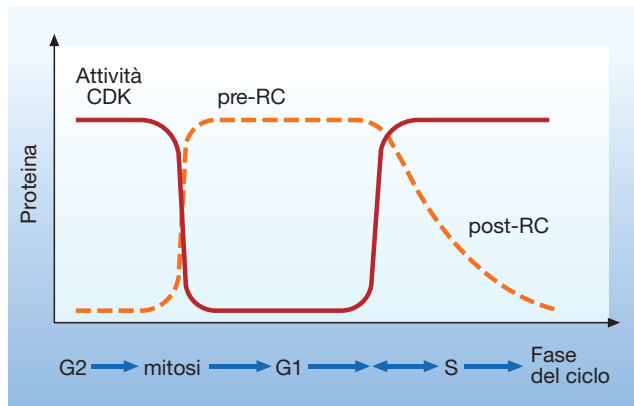


Figura 6.30 Meccanismo di controllo della replicazione del DNA durante il ciclo cellulare. All'uscita dalla mitosi l'attività delle CDK diminuisce drasticamente e rimane bassa per tutta la fase G1 (linea continua). Ciò permette la formazione dei pre-RC (complessi di pre-replicazione) (linea tratteggiata). I pre-RC vengono sostituiti dai post-RC (complessi di post-replicazione) quando l'attività delle CDK risale bruscamente alla transizione tra fase G1 e fase S. I pre-RC possono formarsi nuovamente solo all'uscita dalla mitosi. I rapporti tra formazione dei pre-RC e attività delle CDK assicurano che il DNA possa replicarsi una volta e una volta sola durante ogni ciclo cellulare.

macchina proteica ancora più complessa che diventa competente per la replicazione e innesca la sintesi vera e propria del DNA. Come indicato in **figura 6.29**, eventi di fosforilazione catalizzati dalle CDK provocano il distacco di Cdc6 e Cdt1 e il contemporaneo aggancio di altre proteine (tra cui MCM10, Cdc45, Dpb11 e il complesso GINS). Questi eventi attivano l'elicasi MCM2-7, portando all'apertura dell'origine e al definitivo assemblaggio della macchina replicativa con il caricamento delle polimerasi e di tutte le altre proteine richieste nel processo replicativo. Tale passaggio è osservabile sperimentalmente valutando le dimensioni della macchina proteica replicativa sul DNA mediante esperimenti di ChIP (*Chromatin ImmunoPrecipitation*) e di footprinting *in vivo* (vedi cap. 21). Nel passaggio da pre-RC al complesso di pre-inizio aumentano i componenti e, quindi, anche le dimensioni della macchina proteica, che, a questo stadio, viene definita come **post-RC** (*post-Replication Complex*).

Perché è così importante l'attività delle proteine chinasi CDK nel controllo della replicazione degli eucarioti? Come detto precedentemente, la replicazione del DNA deve essere limitata alla fase S e deve avvenire una volta e una volta soltanto a ogni ciclo cellulare. È proprio il rapporto reciproco tra formazione dei pre-RC e attività delle CDK ad assicurare questo controllo. Infatti, l'attività delle CDK diminuisce quando le cellule escono dalla mitosi e permane a livelli bassi durante tutta la fase G1. Come si può osservare in **figura 6.30**, è soltanto quando l'attività delle CDK è bassa che si possono formare i pre-RC sulle origini di replicazione. Il brusco incremento dell'attività delle CDK all'inizio della fase S innesca la sintesi del DNA permettendo il passaggio dai pre-RC ai post-RC, ma i pre-RC (cioè lo stato competente a replicare il DNA) possono riformarsi solo quando l'attività delle CDK decresce nuovamente, il che può avvenire solo dopo il passaggio della cellula attraverso la mitosi. Questa ingegnosa combinazione permette che il DNA si replichi una sola volta durante ciascun ciclo di divisione cellulare.

6.8 Alterazioni nei meccanismi di replicazione possono provocare cambiamenti quantitativi di sequenze di DNA nel genoma

Nel paragrafo 6.7 abbiamo descritto i meccanismi di controllo della replicazione del genoma che garantiscono che ciascun segmento del genoma sia replicato una volta e una volta sola durante la fase S, in modo tale che le cellule figlie abbiano genomi identici e uguali a quello della cellula materna. In particolare il fatto che negli eucarioti pluricellulari tutte le cellule somatiche hanno genomi uguali

è dimostrato, tra l'altro, dagli esperimenti di John Gurdon (in *Xenopus*) e di Ian Wilmut (pecora Dolly) che ottennero interi individui partendo dal nucleo di cellule differenziate. A questa regola esistono però numerose e interessanti eccezioni. In questo paragrafo descriveremo situazioni in cui uno o più geni sono replicati più di altri in un nucleo cellulare; il maggior numero di copie geniche risulta, naturalmente, in un aumento della sintesi del prodotto genico. In altri casi più rari si ha, al contrario, la perdita dai cromosomi di alcune sequenze genomiche, in genere non codificanti, ma in alcuni casi anche di veri e propri geni che, evidentemente, non sono necessari per il differenziamento della specifica linea cellulare.

Amplificazione genica

Come vedremo più in dettaglio nel paragrafo 12.1, i geni che codificano per l'rRNA sono presenti negli eucarioti in copie multiple ripetute in tandem e formano uno o pochi nucleoli che si trovano nel nucleo cellulare (*vedi* finestra 12.2). Ci sono da 100 ad alcune centinaia di geni per l'rRNA negli eucarioti unicellulari e multicellulari. Questa ridondanza genica è necessaria perché la cellula possa produrre una quantità sufficiente di ribosomi. In particolari condizioni, però, come è il caso degli oociti degli anfibi, la cellula deve sintetizzare ribosomi a una velocità molto maggiore di quanto le è consentito da questo numero di geni. Infatti, l'oocita accumula, durante la sua crescita, una grande quantità di ribosomi che gli permetteranno poi, in una ventina di cicli cellulari molto ravvicinati, un rapido sviluppo fino alla formazione di una larva natante. Ciò è possibile perché il gruppo di geni ripetuti in tandem per l'rRNA subisce un processo di **amplificazione genica**. La **figura 6.31A** mostra come, a livello morfologico, l'amplificazione si presenta come un gran numero di nucleoli, anziché i pochi delle cellule somatiche. Ciascuno di questi nucleoli soprannumerari contiene una lunga molecola di DNA circolare in cui il gene per l'rRNA è ripetuto in tandem centinaia di volte (**fig. 6.31B**). Il meccanismo molecolare alla base di questo processo di amplificazione è ancora in gran parte da chiarire, ma sembra implicare eventi di replicazione extracromosomica combinati a eventi di ricombinazione (*vedi* cap. 8). Un'amplificazione dei geni per l'rRNA analoga a questa è stata osservata anche in altri organismi, per esempio pesci e insetti; in questi ultimi il DNA amplificato forma, anziché tanti nucleoli, un'unica grande massa nucleolare.

Mentre l'amplificazione genica sopra descritta è un processo naturale e regolato, in altri casi l'amplificazione di alcuni geni è il risultato di particolari situazioni patologiche e artificiali. Uno di questi, che non porta però necessariamente alla generazione di copie extra-cromosomiche di tratti genomici, è connesso al fenomeno della resistenza ad agenti chemioterapici spesso usati nella terapia contro il cancro. È stato infatti scoperto che la resistenza al metotrexato (MTX) è causata dalla sovra-produzione del suo bersaglio, l'enzima diidrotolato riduttasi, codificata dal gene DHFR. La conoscenza dei meccanismi che portano all'amplificazione del gene DHFR in cellule resistenti al MTX è oggi frequentemente utilizzata per produrre proteine ricombinanti in cellule di mammifero (*vedi* cap. 20). In generale, l'amplificazione genica o, comunque, un incremento nel numero di copie di alcuni geni, sono frequentemente osservati in alcuni tumori solidi e contribuiscono all'evoluzione del tumore stesso. Tali geni, per lo più, codificano per oncogeni (ad es. MYC) o fattori di crescita, ad esempio l'EGFR o *Epidermal Growth Factor Receptor*.

Politenizzazione

Un caso particolare di amplificazione genica è rappresentato dal fenomeno della politenizzazione, un caso noto da molti anni e che riguarda i cromosomi giganti,

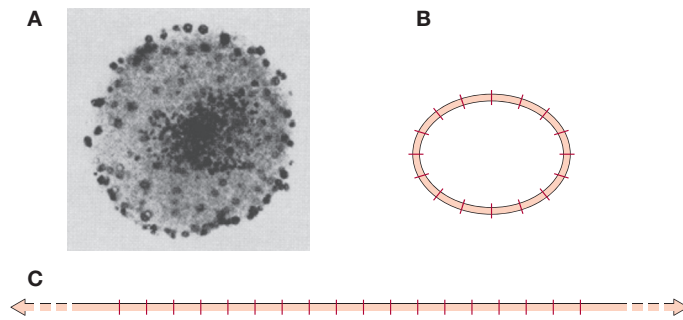


Figura 6.31 Amplificazione dei geni per l'rRNA nell'oocita di *Xenopus*. (A) Fotografia al microscopio ottico del nucleo di un oocita maturo di *Xenopus laevis*. Le piccole masse scure, molto addensate al centro e sparse in tutto il nucleo, sono centinaia di nucleoli extracromosomici. (Fonte: Brown, D.D., David, I.B. (1968), *Science*,

160, pp. 272-280.) (B) Ciascuno dei nucleoli visibili in (A) contiene una lunga molecola di DNA circolare in cui sono ripetute in tandem centinaia di copie del gene per l'rRNA ribosomale, prodotti dell'amplificazione del gruppo di circa 400 copie organizzate in tandem in un locus cromosomico (C).



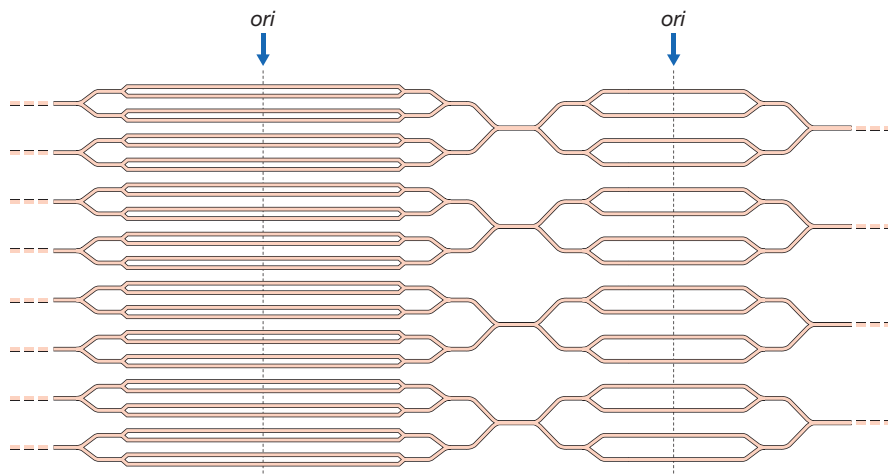
Figura 6.32 Fotografia al microscopio ottico di cromosomi politenici di *Drosophila melanogaster*. Si noti come lo spessore del cromosoma non è costante ma varia nelle diverse regioni. (Fonte: John W. Sedat, UCSF, USA.)

detti appunto politenici, delle ghiandole salivari di *Drosophila*, di cui abbiamo già parlato nel paragrafo 4.5. Nei cromosomi politenici si hanno numerose (fino a circa 1000) molecole di DNA identiche parallele, dovute alla replicazione successiva del DNA cromosomico non seguita dalla separazione delle molecole figlie. Osservando con attenzione la **figura 6.32** si può notare che i cromosomi giganti non hanno uno spessore costante, ma sono più spessi in alcune parti e più sottili in altre, conseguenza del fatto che il numero di repliche subite dalle varie regioni cromosomiche non è uguale lungo tutto il cromosoma. Secondo il modello mostrato in **figura 6.33**, le origini di replicazione vengono attivate più volte, ma non necessariamente lo stesso numero di volte per tutte le origini, risultando così in un diverso grado di politenizzazione per i diversi repliconi; inoltre la replicazione, che procede bidirezionalmente per qualche decina di Kpb, non viene sempre completata fino a confluenza con il replicone adiacente, per cui alcune regioni restano sotto-replicate. Nel caso delle ghiandole salivari la gran parte del genoma è politenizzato, dove più, dove meno.

Questo fenomeno non avviene solo nelle cellule delle ghiandole salivari, ma anche in cellule differenziate di altri tessuti, dove le regioni politenizzate sono in genere di meno e non le stesse. Il risultato di questo differente profilo di politenizzazione in cellule diverse porta alla presenza nei diversi tipi di cellule di un maggiore numero di copie di certe regioni cromosomiche piuttosto che altre e questo ha sicuramente rilevanza per il differenziamento cellulare. Va anche detto

Figura 6.33 Modello per la politenizzazione dei cromosomi.

La politenizzazione è dovuta a multipli inizi di replicazione che avvengono a ciascuna origine (*ori*) localizzata al centro di una regione da politenizzare. Le forche di replicazione che, partendo dalle origini, procedono bidirezionalmente per qualche decina di Kpb, si arrestano prima di arrivare a confluenza con le forche dei repliconi adiacenti.



che la politenizzazione non è limitata agli insetti, ma è abbastanza diffusa in molti organismi, anche nelle piante.

Diminuzione genica, cromatinica e cromosomica

In alcuni rari ma interessanti casi si osserva un fenomeno opposto ai due sopra descritti, cioè la perdita di geni o parti di genoma in alcune cellule somatiche differenziate di un organismo. Questo fenomeno è stato chiamato **diminuzione genica**, **diminuzione cromatinica** o **diminuzione cromosomica**, a seconda che comporti la perdita, rispettivamente, di uno o pochi geni, di estesi tratti cromatinici o di interi cromosomi. Fenomeni di questo genere sono rari ma distribuiti nel mondo animale: sono stati, infatti, descritti in protozoi ciliati, nematodi, crostacei, insetti, ciclostomi e anche marsupiali.

I meccanismi di eliminazione di geni differiscono, tra le diverse specie, per vari aspetti quali lo stadio di sviluppo in cui avvengono, il tipo di cellula e di tessuto, lo stadio del ciclo cellulare, la quantità di materiale eliminato ecc. Storicamente, la prima descrizione di diminuzione cromosomica è stata fatta da Theodor Boveri oltre cento anni fa nel nematode *Ascaris equorum*. Successivamente, il fenomeno è stato studiato un po' più a fondo, anche a livello molecolare, soprattutto nel nematode *Ascaris suum*, parassita intestinale del maiale e in alcune specie di Copepodi (microcrostacei planctonici). Naturalmente, la diminuzione genica può riguardare alcune o tutte le cellule somatiche, ma mai le cellule della linea germinale, che devono, per definizione, mantenere intatto l'intero patrimonio genetico. I meccanismi, abbastanza vari nei diversi organismi, non sono casuali, ma sempre ben regolati e controllati. In alcune situazioni si ha una frammentazione dei cromosomi, seguita dall'aggiunta da parte della telomerasi di nuovi telomeri alle estremità dei frammenti cromosomici da conservare, mentre gli altri frammenti vanno perduti. In altri casi sembra esservi escissione di regioni eterocromatiniche interne ai cromosomi, che vanno perdute. Le regioni cromatiniche eliminate contengono soprattutto sequenze genomiche ripetute ed eterocromatiniche, ma, in alcuni casi, è stata descritta la perdita anche di qualche gene che, evidentemente, non è necessario in quella particolare cellula differenziata. I meccanismi alla base di questi processi non sono stati ancora ben chiariti.

Micro- e macronucleo dei protozoi ciliati

Una situazione complessa e insolita è presente nei protozoi ciliati. Questi sono dotati di due nuclei: un micronucleo piccolo, diploide e trascrizionalmente inat-

tivo, che ha il ruolo di conservare e trasmettere alla progenie il genoma intatto; un macronucleo vegetativo, trascrizionalmente attivo, implicato nella sintesi degli RNA e quindi delle proteine cellulari. Il genoma contenuto nel macronucleo è molto particolare in quanto costituito da un grandissimo numero di molecole di DNA, corrispondenti a frammenti genomici, contenenti uno o pochi geni; ce ne sono migliaia di copie per ciascuno dei 10-20 000 geni diversi.

Durante le normali duplicazioni cellulari il micronucleo si divide per mitosi trasmettendo così sempre un genoma completo; il macronucleo invece si divide per “amitosi”, un processo in cui la replicazione dei frammenti monogenici di DNA è seguita da una loro ripartizione non regolata tra i nuclei figli. Come abbiamo detto, queste copie geniche sono trascrizionalmente attive e responsabili della sintesi delle proteine della cellula.

Dopo circa 200 generazioni, però, le cellule figlie mostrano segni di invecchiamento, in parte conseguenza di un progressivo sbilanciamento del numero di copie dei diversi geni nel macronucleo. Questa situazione viene superata attraverso un ciclo di riproduzione sessuale, che comporta una meiosi, una coniugazione tra due cellule che si scambiano materiale genetico e, infine, la ricostituzione di un nuovo macronucleo a partire da un micronucleo. A tal fine, il micronucleo va incontro a una serie di replicazioni non seguite da separazione delle molecole di DNA replicate, analogamente alla politenizzazione che abbiamo già visto avvenire nelle ghiandole salivari di *Drosophila* (fig. 6.32). Successivamente, questi cromosomi politenici subiscono una frammentazione cui segue la perdita di gran parte delle regioni intergeniche non codificanti del genoma e la produzione di frammenti relativamente piccoli di DNA, ciascuno contenente un gene con le regioni fiancheggianti necessarie per l'espressione. Nel complesso, quindi, la formazione del macronucleo da un micronucleo è il risultato di una politenizzazione dei cromosomi seguita da frammentazione e diminuzione cromatinica.

Lecture di approfondimento

- DePamphilis, M.L., Blow, J.J., Ghosh, S., Saha, T., Noguchi, K. e Vassilev, A. (2006), «Regulating the Licensing of DNA Replication Origins in Metazoa», *Current Opinion in Cell Biology*, **18**, pp. 231-239.
- Duharcourt, S., Lepère, G. e Meyer, E. (2009), «Developmental genome rearrangements in ciliates: a natural genomic subtraction mediated by non-coding transcripts», *Trends in Genetics*, **25**, pp. 344-350.
- Meselson, M. and Stahl, F.W. (1958), «The replication of DNA in *Escherichia coli*», *Proc. Natl. Acad. Sci. USA*, **44**, pp. 671-682.
- Moldovan, G.L., Pfander, B. e Jentsch, S. (2007), «PCNA, the Maestro of the Replication Fork», *Cell*, **129**, pp. 665-679.
- Pomerantz, R.T. e O'Donnell, M. (2007), «Replisome Mechanics: Insights into a Twin DNA Polymerase Machine», *Trends in Microbiology*, **15**, pp. 156-164.

7

Riparazione del DNA

7.1 Mutazioni

7.2 Sistemi di riparazione del DNA

7.3 Risposta cellulare a danni sul DNA e connessioni tra riparazione e ciclo cellulare

Il mantenimento della corretta sequenza nucleotidica del DNA è di fondamentale importanza per il suo ruolo di molecola depositaria dell'informazione genetica di una cellula. Proprio per permettere che l'informazione genetica sia trasmessa pressoché inalterata da una generazione all'altra, la replicazione del DNA (*vedi* cap. 6) è un processo molto preciso.

Tuttavia, contrariamente a quanto era stato inizialmente ipotizzato, la struttura del DNA è intrinsecamente molto fragile. A 37 °C (la temperatura a cui si trovano la maggior parte delle cellule umane) circa 18 000 basi azotate al giorno sono perse da ogni cellula, a causa della rottura del legame glicosidico che le collega con la molecola dello zucchero deossiribosio. Ogni cellula umana, a temperatura e pH fisiologico, subisce, in 24 ore, 100-500 eventi di deaminazione che causano la trasformazione di una citosina in uracile. Inoltre, una serie di reazioni biochimiche necessarie per il metabolismo stesso della cellula producono specie reattive dell'ossigeno (ROS) che alterano l'integrità del DNA.

Gli esempi sopra indicati sono soltanto alcuni dei possibili danni che, in condizioni fisiologiche (danni intrinseci), possono alterare la struttura del DNA. Ma il DNA è continuamente soggetto anche all'attacco di agenti chimico-fisici esterni che possono causare una serie di danni estrinseci sul DNA. Per esempio, le radiazioni ultraviolette presenti nella luce solare producono sul DNA una serie di fotoprodotto che determinano una distorsione della doppia elica. Radiazioni ionizzanti presenti nei raggi cosmici, o generate da prodotti radioattivi naturali o da tecnologie utilizzate nella pratica medica, possono provocare rotture del legame fosfodiesterico su uno o entrambi i filamenti del DNA. Infine, una serie di composti chimici utilizzati in lavorazioni industriali, presenti nel fumo di sigaretta o anche in componenti della nostra dieta quotidiana, possono causare lesioni e danni sul DNA.

Prima di discutere i meccanismi molecolari che controllano la stabilità del genoma è necessario ricordare brevemente (si rimanda a un testo di Genetica per una trattazione più ampia) quali siano le conseguenze genetiche di alterazioni nella sequenza del DNA causate da errori nel processo di replicazione e da danni al DNA spontanei o indotti.

7.1 Mutazioni

Variazioni nella sequenza nucleotidica del DNA prendono il nome di **mutazioni**. La maggior parte delle mutazioni sono definite **puntiformi**, in quanto

determinano il cambiamento di un singolo nucleotide. Le mutazioni puntiformi sono divise in due categorie: le **transizioni**, che sono cambiamenti da una purina a un'altra o da una pirimidina a un'altra, e le **trasversioni**, che sono cambiamenti da purina a pirimidina o da pirimidina a purina.

Come mostrato in **figura 7.1A**, se durante il processo replicativo si introduce un appaiamento non corretto tra i due filamenti del DNA e tale errore non viene riparato, al secondo ciclo di replicazione del DNA la mutazione viene fissata nel genoma. L'effetto fenotipico di una mutazione puntiforme può essere molto diverso. Se il cambiamento nucleotidico cade nella regione di un gene codificante per una proteina ci sono tre possibilità (**fig. 7.1B**):

1. la mutazione cambia la sequenza di una tripletta (codone) codificante per un certo amminoacido, ma, in virtù della degenerazione del codice genetico, il nuovo codone codifica per lo stesso amminoacido codificato dal codone originario: in tal caso la mutazione è definita **silente** o **sinonima** perché a livello della proteina codificata dal gene non si determina alcun cambiamento;
2. la mutazione puntiforme in regioni codificanti per una proteina può risultare nella formazione di un codone con un significato diverso dall'originario, con conseguente cambiamento di un singolo amminoacido: tali mutazioni vengo-

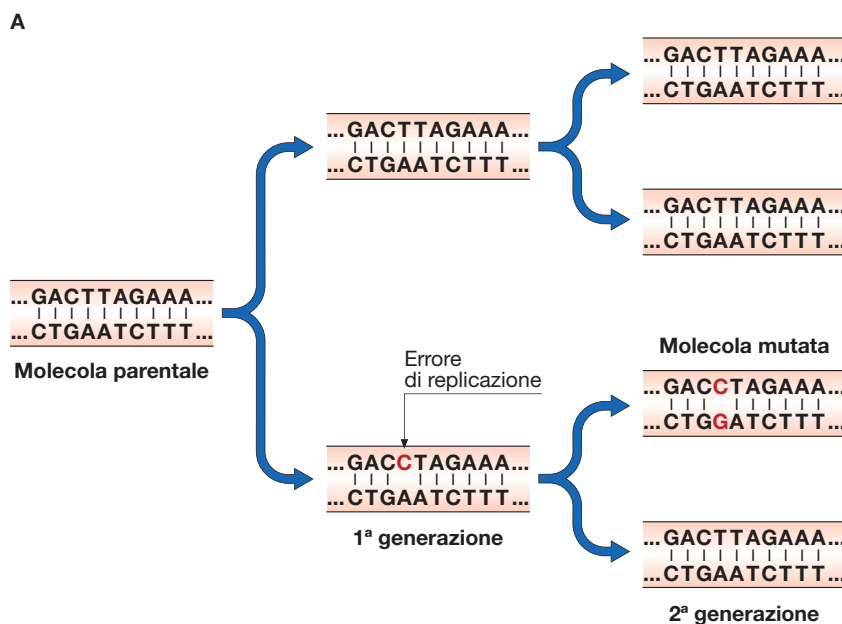


Figura 7.1 Esempi di mutazioni. (A) Durante la replicazione possono avvenire rari errori che determinano un appaiamento sbagliato ("mismatch") in una delle due eliche figlie (1ª generazione). Se il mismatch non viene riparato dal meccanismo appropriato ("mismatch repair"), la molecola portatrice del mismatch fisserà la mutazione in una delle due molecole figlie di 2ª generazione. (B) Mutazioni puntiformi per sostituzione di una base in una regione codificante per una proteina possono generare effetti diversi: (I) mutazione silente (non provoca alcun cambiamento di amminoacido); (II) mutazione missenso (provoca il cambiamento di un amminoacido); (III) mutazione non senso (provoca la formazione di una tripletta non senso con conseguente blocco della sintesi proteica); (IV) l'inserzione o la delezione di un singolo nucleotide nella regione codificante per una proteina danno origine a una mutazione frameshift, in cui il modulo di lettura delle triplette è totalmente alterato dal punto della mutazione in poi.

B

I) Mutazione silente UAU → UAC Tyr Tyr II) Mutazione missenso UAU → UCU Tyr Ser III) Mutazione non senso UAU → UAA Tyr Stop	IV) Mutazione frameshift Sequenza normale delle basi: ATG AAA GAG UAU Met Lys Glu Tyr Aggiunta di una base: ATG AAG AGA GUA U Met Lys Arg Val Delezione di una base (manca una A): ATG AAG AGU AU Met Lys Ser
--	--

no chiamate **missenso** e la mutazione può alterare le proprietà della proteina determinando una variazione nel fenotipo;

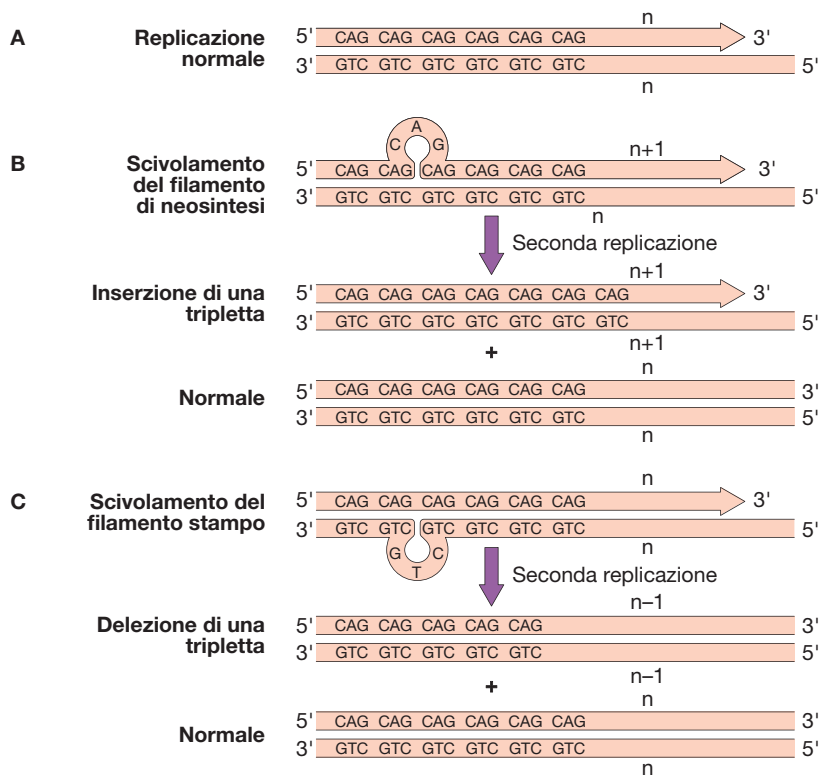
- la sostituzione di un singolo nucleotide nella regione codificante di un gene può determinare la formazione di una delle 3 triplette non senso del codice genetico: tali mutazioni sono chiamate **non senso** e causano la formazione di una proteina tronca dovuta a terminazione prematura della sintesi proteica e, quindi, un probabile fenotipo mutante.

Nel DNA possono verificarsi, anche se con frequenza molto minore, delle mutazioni per sostituzione, delle **delezioni** o **inserzioni** di un singolo nucleotide. Questa perdita o aggiunta di un nucleotide determina uno sfasamento del codice di lettura di un gene codificante per una proteina. Queste mutazioni vengono chiamate **frameshift** e alterano tutta la sequenza di amminoacidi a valle del punto in cui è avvenuta la mutazione (**fig. 7.1B**). Anche nel caso delle mutazioni frameshift è molto probabile che si generi un fenotipo mutante, poiché la sequenza amminoacidica della proteina è fortemente alterata.

Alcune regioni del DNA hanno un'aumentata instabilità genomica a causa della ripetizione di sequenze nucleotidiche. Se una sequenza di pochi nucleotidi si ripete più volte in una regione del DNA, durante il processo replicativo, le sequenze ripetute possono formare appaiamenti strutturali alternativi con una stabilità termodinamica simile a quella dell'appaiamento corretto. Le polimerasi possono, quindi, "scivolare" durante la replicazione di queste regioni, portando a una contrazione o a un'espansione delle sequenze ripetute stesse (**fig. 7.2**). Nell'uomo è stato trovato che alcune malattie neurologiche (ad es. la corea di Huntington, la sindrome dell'X fragile, la distrofia miotonica e altre) sono dovute proprio a un'espansione di triplette in specifiche regioni del genoma.

Gli effetti generali di mutazioni sono ampiamente descritti nei testi di Genetica. Qui vogliamo ricordare soltanto alcuni aspetti generali importanti per gli orga-

Figura 7.2 Espansione o contrazione di triplette durante la replicazione. Alcune regioni sul DNA contengono numerose copie di triplette uguali disposte in serie. La replicazione di tali regioni normalmente non provoca variazioni nel numero di triplette (A). Tuttavia, può accadere che durante la replicazione si formino degli appaiamenti intramolecolari (che si suppone siano dovuti a degli scivolamenti dell'apparato replicativo in tali regioni) sul filamento di neosintesi (B) o sul filamento stampo (C). Come conseguenza, alla seconda generazione, si genera nel caso (B) l'inserzione di una o più triplette e nel caso (C) la delezione di una o più triplette.



nismi eucariotici diploidi pluricellulari, tra cui l'uomo. Innanzitutto, una certa mutazione può avvenire in una cellula somatica o in una cellula germinale. Chiaramente le conseguenze della mutazione sono molto diverse perché una cellula somatica è importante solo per l'organismo in cui si trova. Se la mutazione è dannosa per quella specifica cellula, di solito, ciò risulta nella morte della cellula stessa con conseguenze per lo più insignificanti sull'intero organismo. Se, tuttavia, come vedremo più avanti, una mutazione influenza specificamente il controllo della stabilità del genoma, può accadere che una mutazione faciliti l'insorgenza di altre mutazioni provocando una reazione a catena che può essere dannosa per l'intero organismo, come accade nei meccanismi di cancerogenesi. Mutazioni in cellule germinali sono molto più rilevanti perché vengono trasmesse alla generazione successiva e saranno presenti in tutte le cellule dell'individuo che eredita la mutazione.

La maggior parte delle mutazioni, incluse tutte quelle silenti e la maggior parte di quelle missenso, non influenzerà il fenotipo dell'organismo in modo significativo. Le mutazioni che causano un fenotipo possono essere divise in due categorie.

1. La perdita di funzione o **loss of function** è la conseguenza di una mutazione che abolisce o riduce l'attività di una proteina. Per lo più, le mutazioni loss of function sono recessive e, quindi, in un organismo eterozigote che porta anche una versione normale del gene l'effetto della mutazione recessiva è mascherato dalla presenza dell'allele normale. In alcuni casi, tuttavia, la mutazione può determinare un fenotipo: si parla di **aploinsufficienza** se l'organismo eterozigote non è in grado di tollerare la riduzione al 50% del livello di proteina normale.
2. La mutazione ad acquisizione di funzione o **gain of function** è piuttosto rara e si verifica allorché la mutazione è in grado di conferire alla proteina alterata codificata dal gene mutato una funzione anomala. Può anche accadere che mutazioni gain of function si verifichino in regioni del genoma che non codificano per proteine, ma che controllano l'espressione di geni. Considerata la loro natura, le mutazioni gain of function sono solitamente dominanti.

7.2 Sistemi di riparazione del DNA

Molecole alterate e mal funzionanti vengono normalmente eliminate dalla cellula. Questo non può avvenire per molecole di DNA danneggiate in quanto ciò comporterebbe la morte stessa della cellula dato che il DNA è il depositario dell'informazione genetica. Il DNA è, quindi, l'unica molecola per cui è stato necessario che nel corso dell'evoluzione si sviluppassero meccanismi deputati alla riparazione.

Poiché le lesioni intrinseche ed estrinseche che si formano sul DNA sono molto diverse tra di loro, le cellule hanno sviluppato numerosi e diversificati sistemi di riparazione. Sarebbe troppo complesso fornire una descrizione dettagliata di tutti i tipi di alterazioni possibili sul DNA. Comunque, per difendersi da questa pletora di lesioni tutti gli organismi, sia procarioti che eucarioti, hanno sviluppato sistemi di riparazione molto diversi tra loro e specializzati nel rimuovere particolari modificazioni sul DNA. Alcuni meccanismi di riparazione sono molto precisi, altri meno, ma tutti hanno lo scopo di permettere la sopravvivenza della cellula in condizioni di danno al DNA. I processi di riparazione sono diversi, non solo in funzione del tipo di danno, ma anche in relazione alla fase del ciclo cellulare (*vedi oltre*) in cui il danno viene prodotto.

In questo testo descriveremo, in un certo dettaglio, i meccanismi molecolari dei più rilevanti sistemi di riparazione negli eucarioti, e in particolare nell'uomo, dato che il cattivo funzionamento di tali meccanismi causa una serie di patologie specifiche e, in generale, una maggior predisposizione all'insorgenza di tumori. Va però detto che gli stessi meccanismi di riparazione sono presenti anche nei procarioti (**finestra 7.1**), a indicare la loro importanza e conservazione nel corso dell'evoluzione.

FINESTRA 7.1

I MECCANISMI DI RIPARAZIONE SONO FUNZIONALMENTE CONSERVATI IN TUTTI GLI ORGANISMI VIVENTI, DAI PROCARIOTI AGLI EUCARIOTI

L'integrità del genoma è sotto costante attacco da parte di agenti genotossici sia nei procarioti che negli eucarioti. Nel testo abbiamo descritto i principali meccanismi di riparazione del DNA negli eucarioti, con particolare riferimento ai processi riparativi presenti nelle cellule umane. Questo perché alterazioni nei meccanismi di riparazione nell'uomo sono causa di gravi patologie e, in particolare, dell'insorgenza di tumori. È però importante ricordare che anche i batteri posseggono sistemi di riparazione del DNA e, come spesso in Biologia Molecolare, lo studio di tali meccanismi in organismi semplici come i procarioti è stato di estrema importanza per individuare analoghi sistemi negli organismi superiori.

Per ragioni di spazio non possiamo qui entrare nei dettagli dei meccanismi riparativi nei procarioti. Pertanto, prenderemo in considerazione solo la riparazione del DNA in cellule di *Escherichia coli*, ricordando, però, che esistono differenze anche tra specie batteriche diverse; tratteremo, poi, i meccanismi che presentano differenze sostanziali a livello molecolare con quanto già descritto nel testo relativamente agli eucarioti. In molti casi, i meccanismi molecolari sono simili e cambia solo la nomenclatura delle proteine coinvolte.

A livello evolutivo i processi riparativi possono essere raggruppati essenzialmente in 3 classi:

1. riparazione diretta del danno, in cui le lesioni vengono revertite chimicamente;
2. riparazione tramite escissione, in cui segmenti più o meno lunghi di un filamento di DNA danneggiato vengono eliminati con successiva sintesi riparativa utilizzando il filamento intatto come stampo;
3. riparazione tramite ricombinazione, in cui un tratto di DNA intatto è utilizzato per riparare l'omologo danneggiato tramite ricombinazione.

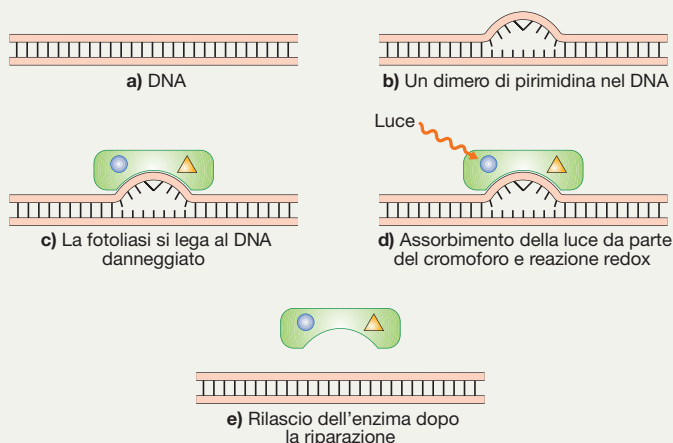


Figura 7F1.1 Reazione di fotoriattivazione di lesioni causate da radiazioni UV. Illustrazione schematica della reazione catalizzata dalla fotoliasi per riparare un dimer di pirimidina. I due simboli colorati all'interno del disegno indicano i due gruppi cromofori dell'enzima.

La **riparazione diretta del danno** è presente sia in cellule di *E. coli* che nel lievito e in alcuni altri eucarioti. I dimeri di pirimidine e i (6-4) pirimidina-pirimidone fotoprodotto (6-4 PP), che sono le lesioni più frequenti causate da radiazioni UV, vengono principalmente rimossi dal NER (*vedi* testo). Tuttavia, entrambe le lesioni sopra citate possono essere rimosse dal DNA mediante una reazione chiamata **fotoriattivazione enzimatica** (**fig. 7F1.1**). La scoperta del ruolo della luce nel favorire la sopravvivenza di batteriofagi o spore fungine in seguito al trattamento con raggi UV avvenne ancor prima della determinazione della struttura del DNA, ma fu solo negli anni sessanta del secolo scorso che si scoprì il meccanismo enzimatico della reazione. L'enzima in grado di revertire i fotoprodotto causati da raggi UV è genericamente chiamato **fotoliasi**. Questo enzima è una flavoproteina che contiene due gruppi cromofori: quando luce di lunghezza d'onda compresa fra 320 e 370 nm colpisce la proteina, scatta la reazione enzimatica che prevede la rottura del dimero e la ricostituzione della struttura corretta delle basi (fotoriattivazione). La reazione procede attraverso il trasferimento di elettroni e nella reazione redox la molecola FAD agisce da donatore di elettroni, mentre il dimero agisce da accettore.

I meccanismi di **escissione delle basi (BER)** o di **escissione di nucleotidi (NER)** sono presenti in tutte le specie batteriche. La logica funzionale dei due processi è sostanzialmente identica nei procarioti e negli eucarioti: cambiano i nomi delle proteine coinvolte e alcune peculiarità dei loro meccanismi molecolari (rimandiamo a testi specialistici).

Il **mismatch repair** o **MMR** che si prende cura degli errori che possono accadere durante il processo di replicazione del DNA è presente in *E. coli* ed è anche quello più conosciuto a livello molecolare (schematizzato in **figura 7F1.2**). L'errore di appaiamento (mismatch) induce una distorsione della doppia elica riconosciuta dall'omodimero MutS che si lega in corrispondenza

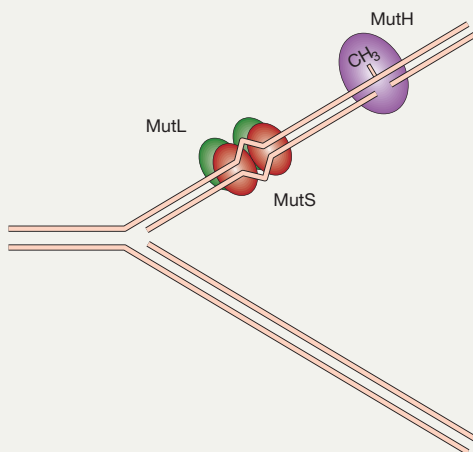


Figura 7F1.2 Riparazione di un errore replicativo tramite il mismatch repair (MMR) nel batterio *Escherichia coli*. *Vedi* testo per i dettagli.

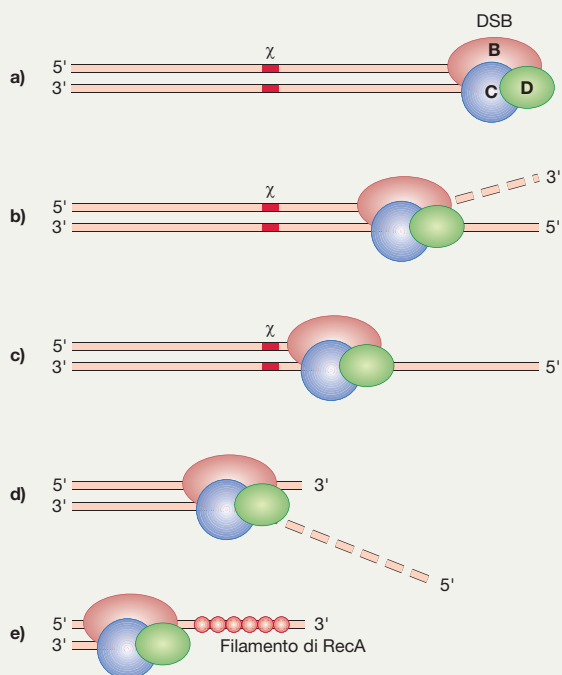


Figura 7F1.3 Azione del complesso RecBCD nel processamento di un DSB (Double Strand Break) in *E. coli*. Il complesso RecBCD si lega a un DSB e srotola la doppia elica tramite la sua attività elicasi in modo ATP-dipendente. L'estremità 3' è rapidamente degradata dalla sua attività nucleasica. Quando, però, il complesso proteico incontra una particolare sequenza di 8 pb sul DNA di *E. coli*, chiamata χ , l'attività nucleasica comincia ad agire sull'altro filamento, generando, alla fine, un filamento sporgente in 3' su cui si lega la proteina RecA che innesci i successivi passaggi ricombinativi.

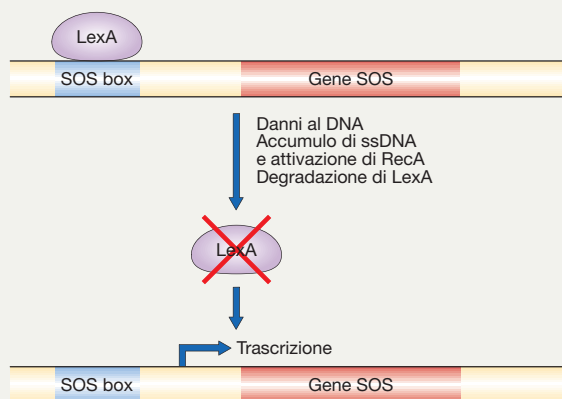


Figura 7F1.4 Il sistema SOS in *E. coli* agisce sul repressore trascrizionale LexA. La proteina LexA si lega ai SOS box presenti nell'operatore dei geni SOS la cui espressione è indotta in risposta a danni al DNA. Come descritto nel testo, la presenza di danni nel DNA determina l'attivazione della proteina RecA, che, legandosi a LexA, induce la sua autoproteolisi determinando l'attivazione dei geni SOS.

za del mismatch; tale legame è stabilizzato dall'interazione con l'omodimero MutL. Un aspetto importante nel MMR è che il filamento di DNA parentale deve essere distinto da quello di neosintesi, così che la riparazione del mismatch avvenga su quest'ultimo. In *E. coli* tale distinzione si basa sullo stato di me-

tilazione del DNA neo-sintetizzato. L'adenina delle sequenze GATC del DNA di *E. coli* è normalmente metilata ma, dopo la replicazione, il filamento neosintetizzato rimane non metilato per una breve finestra temporale. Il fatto che il filamento parentale sia metilato, mentre quello di neosintesi non lo è, permette di distinguere quale filamento contenga l'errore replicativo e debba, quindi, essere riparato. Il complesso MutS-MutL attiva la proteina MutH che si lega alla sequenza GATC metilata più vicina e, in funzione della sua attività endonucleasica, taglia il filamento di neosintesi di fronte alla base metilata. L'escissione di un tratto di DNA contenente il mismatch è realizzata tramite l'azione combinata di un'elicasi che srotola parzialmente il filamento contenente il mismatch e la successiva azione di una esonucleasi con la corretta polarità. Il DNA è poi riparato tramite l'azione coordinata della DNA polimerasi III e della DNA ligasi.

La **riparazione per ricombinazione omologa (HR)** è molto attiva in *E. coli* e si realizza utilizzando una varietà di sotto-pathway molto complessi per cui, ancora una volta, rimandiamo a testi o articoli specialistici. La ricombinazione omologa è innescata da danni al DNA, quali rotture a doppio filamento (DSB), o da processi fisiologici, quale la coniugazione batterica. La logica funzionale del meccanismo di ricombinazione omologa nei procarioti è analoga a quella descritta nel testo relativamente agli eucarioti. Per iniziare la ricombinazione a un DSB, per esempio, il DNA deve essere anche in *E. coli* modificato e il complesso **RecBCD** gioca un ruolo essenziale dato che agisce da una parte come elicasi, aprendo il DNA, e dall'altra come nucleasi, producendo alla fine un tratto di DNA sporgente in 3' che viene ricoperto dalla proteina **RecA** (fig. 7F1.3). RecA, in *E. coli*, è la proteina chiave richiesta essenzialmente in tutti i sotto-pathway ricombinativi per invadere la molecola di DNA omologa e per promuovere l'appaiamento con il filamento complementare. Risultati molto recenti suggeriscono che anche in *E. coli* possa esistere un meccanismo di NHEJ analogo a quello descritto negli eucarioti.

Il **sistema SOS** è il principale sistema di *Post-Replication Repair* (PRR) in *E. coli* e coinvolge la proteina RecA descritta sopra. Durante la crescita in condizioni normali, l'espressione di geni chiamati SOS o *din* (*damage inducible*) è repressa tramite il legame di una proteina repressore, chiamata **LexA**, su una sequenza nota come "SOS box" presente sul sito operatore di tali geni. Come conseguenza a danni al DNA si accumulano regioni di DNA a singolo filamento (ssDNA): RecA ha una grande affinità per il ssDNA e il legame a tali regioni determina la sua attivazione. La forma attivata di RecA è in grado di legarsi al repressore LexA e di indurre la sua distruzione tramite l'attività proteolitica che LexA stessa possiede (fig. 7F1.4). Questo causa la derepressione, cioè l'attivazione, di una serie di geni i cui prodotti sono richiesti per una corretta risposta a danni al DNA.

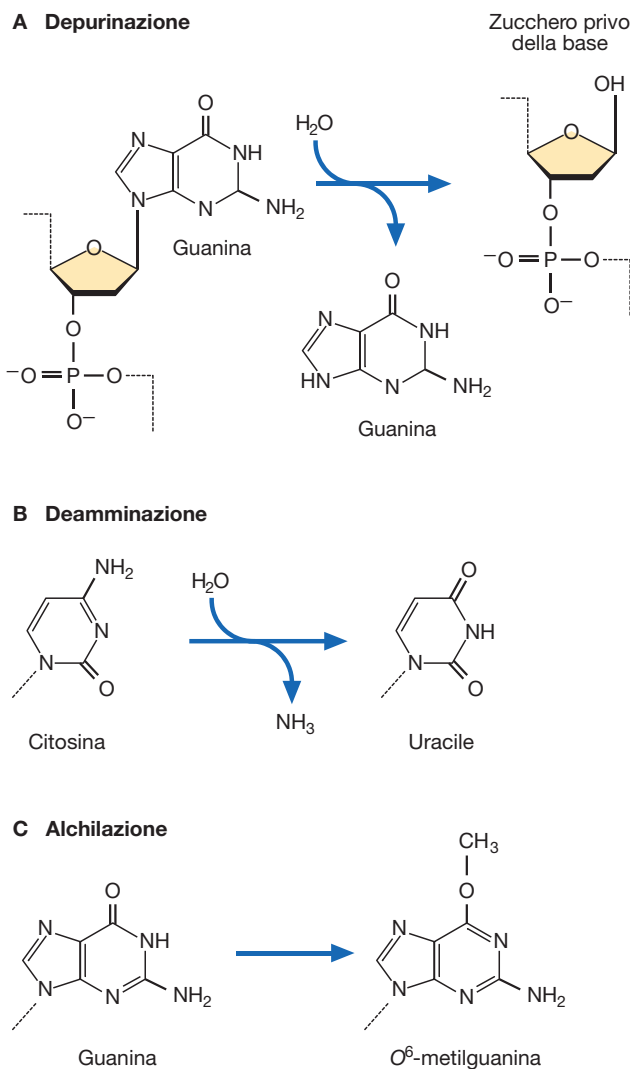
Un analogo sistema SOS non è presente negli eucarioti: tuttavia, la trascrizione di molti geni eucariotici è indotta dall'attivazione del checkpoint innescato da danni al DNA, che è, quindi, il processo più simile al sistema SOS di *E. coli*.

Riparazione per escissione delle basi (BER)

I danni al DNA di origine endogena (**fig. 7.3**) causati da specie reattive dell'ossigeno, ma anche l'idrolisi delle basi azotate del DNA, la perdita spontanea di gruppi amminici da alcune di esse (C, A, G) e altri, sono riparati attraverso il sistema di **riparazione per escissione delle basi**, denominato **BER** (*Base Excision Repair*). Per avere un'idea del numero di lesioni prodotte dal normale metabolismo cellulare, basti ricordare che i composti reattivi dell'ossigeno possono causare un centinaio di diverse modificazioni ossidative sul DNA che, nel loro insieme, possono determinare 2×10^4 lesioni al giorno in una cellula umana.

L'importanza del BER è dimostrata da esperimenti di knock-out in topo (*vedi cap. 23*): in tali tipi di esperimenti, con opportune tecniche genetiche si inattiva la funzionalità del gene in esame. Si ottengono così dei modelli murini che permettono di valutare l'importanza e, talvolta, la possibile funzione del gene che è stato inattivato. Se un gene del BER viene spento in tali modelli animali, l'embrione di topo non riesce a svilupparsi, evidenziando l'importanza del BER per la vitalità cellulare e per il completamento dello sviluppo intrauterino dell'animale. Poiché, in tali condizioni, l'embrione non è normalmente soggetto all'azione di agenti chimico-fisici esterni, l'incapacità di completare lo sviluppo

Figura 7.3 Modificazioni di nucleotidi causate da danni endogeni al DNA. Le modificazioni di nucleotidi più comuni causate da danni endogeni al DNA, in particolare da radicali liberi dell'ossigeno, sono: (A) la depurinazione idrolitica di nucleotidi purinici; (B) la deaminazione idrolitica di basi azotate contenenti gruppi amminici; (C) l'alchilazione di varie posizioni sulle basi azotate.



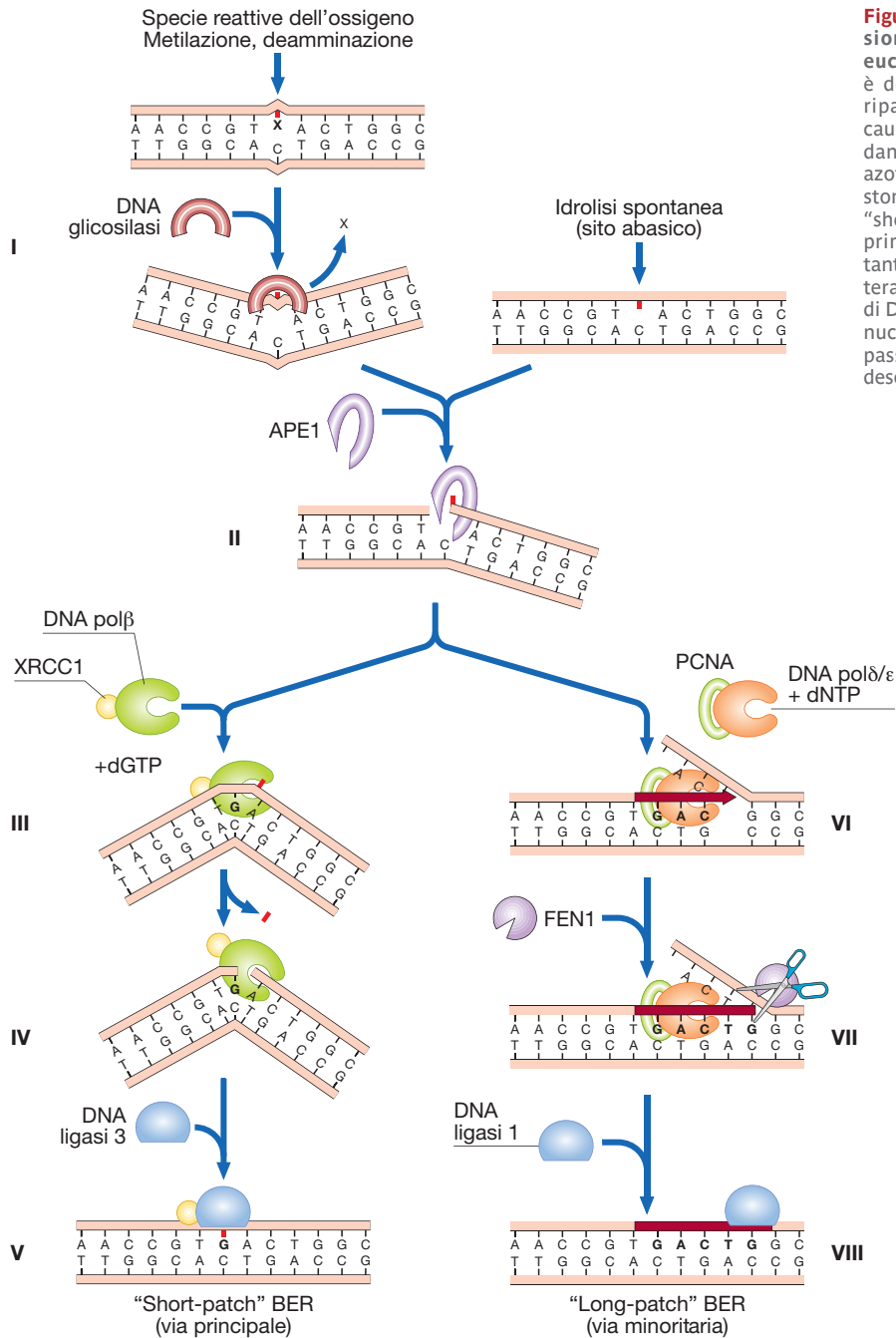


Figura 7.4 Riparazione per escissione delle basi (BER) negli eucarioti. Questo meccanismo è deputato principalmente alla riparazione di danni endogeni causati da stress ossidativo o da danni esogeni causati sulle basi azotate da agenti mutageni. Esistono due vie del BER, chiamate "short-patch" o "long-patch": nel primo caso viene sostituito soltanto il nucleotide con la base alterata (X), nel secondo la regione di DNA riparata è di una decina di nucleotidi. Il dettaglio dei diversi passaggi nelle due vie del BER è descritto nel testo.

quando il BER è inattivato evidenzia il ruolo di tale meccanismo di riparazione nella rimozione di lesioni intrinseche causate dal normale metabolismo cellulare.

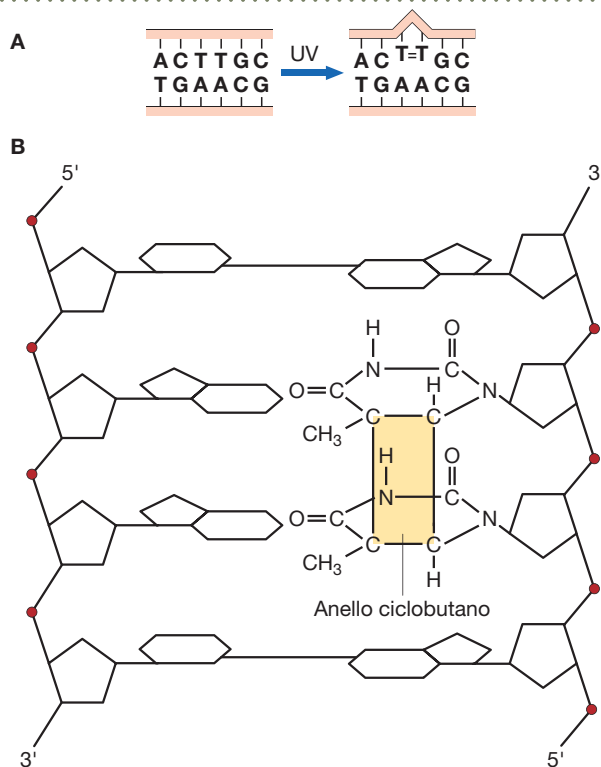
Come mostrato in **figura 7.4**, esistono due modalità di esecuzione del BER. Il primo passaggio del processo è il riconoscimento del danno da riparare. Nel caso di danni causati da specie reattive dell'ossigeno (metilazione o deaminazione), **DNA glicosilasi** specifiche, in grado di riconoscere i vari tipi di lesioni, rompono il legame glicosidico tra lo zucchero e la base azotata danneggiata. Questi enzimi, identificata la base alterata, comprimono la doppia elica del DNA così che la base alterata viene inglobata in una cavità presente nella struttura tridimensionale della glicosilasi. L'enzima taglia poi il legame glicosidico che lega la base azotata

alterata al deossiribosio sul DNA (fase I a sinistra nella figura). Si crea così un sito abasico (chiamato sito AP) che si può produrre anche per idrolisi spontanea della base (fase I a destra nella figura). Successivamente, una endonucleasi chiamata APE1 riconosce il sito privo della base e incide il legame fosfodiesterico (fase II). A questo punto il meccanismo del BER può procedere attraverso due vie, la prima delle quali è preponderante nei mammiferi e viene chiamata **short-patch BER**. La DNA polimerasi β , attraverso l'attività liasica che tale polimerasi possiede, rimuove il deossiribosio privo della base e la stessa pol β riempie il buco di un nucleotide così creatosi (fase III a sinistra). Successivamente, il legame fosfodiesterico è saldato da una ligasi specializzata (DNA ligasi 3) con l'aiuto di una proteina nota come XRCC1 (fase V a sinistra). La seconda via (indicata a destra come **long-patch BER**) prevede l'azione delle DNA polimerasi δ o ϵ e di PCNA che allungano il 3'OH di 2-10 nucleotidi, attraverso una reazione di "strand displacement" (fase VI a destra). Il filamento di DNA spelato via è poi tagliato da una nucleasi, denominata FEN1 (*Flap Endonuclease 1*) (fase VII a destra), che riconosce la particolare struttura che si forma durante la reazione. La discontinuità sul DNA è poi saldata dall'azione della DNA ligasi 1 (fase VIII a destra).

Riparazione per escissione di nucleotidi (NER)

La **riparazione per escissione di nucleotidi** o **NER** (*Nucleotide Excision Repair*) è coinvolta nella riparazione di una varietà di lesioni che provocano una distorsione della doppia elica del DNA e sono causate da agenti chimico-fisici. In particolare, il NER è richiesto per la riparazione dei danni causati da radiazioni ultraviolette (UV) che sono presenti nella normale luce solare. Basti pensare che è stato calcolato che, giornalmente, in ciascuna cellula della pelle esposta alla luce del sole a un'altitudine di 600 metri vengono prodotti decine di migliaia di fotoprodotti causati da radiazioni UV, i più comuni dei quali sono i dimeri di pirimidina (**fig. 7.5**). Molti

Figura 7.5 I dimeri di pirimidina sono tra i danni più comuni causati dalle radiazioni UV. (A) Le radiazioni UV determinano la formazione di legami covalenti tra due pirimidine (T o C) sullo stesso filamento che provocano una distorsione della doppia elica. (B) Dettaglio di un dimeri tra due residui di timina con l'anello di ciclobutano che si forma in evidenza.



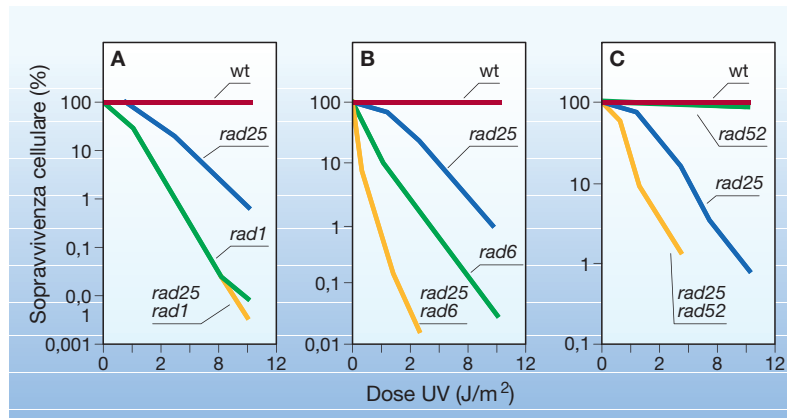


Figura 7.6 Identificazione di geni di lievito codificanti per proteine coinvolte nella riparazione di danni al DNA e definizione di gruppi di epistasi. Cellule di lievito wild-type (wt) sono quasi totalmente resistenti a radiazioni UV fino a dosi intorno a 10 J/m². Mutanti di lievito alterati nel meccanismo di riparazione di lesioni causate da raggi UV sono più sensibili di cellule wt e appartengono alla categoria dei mutanti definiti *rad* (*radiation sensitive*). (A) Sono mostrate le curve di sopravvivenza in risposta a diverse dosi di radiazioni UV dei mutanti *rad25* e *rad1*. La sensibilità alle radiazioni dei due mutanti è diversa, ma il doppio mutante *rad1 rad25* ha la stessa sensibilità del mutante singolo più sensibile (*rad1*): entrambi i mutanti sono coinvolti in funzioni richieste per il NER

e appartengono, pertanto, allo stesso gruppo di epistasi (*vedi* testo). (B e C) *rad6* e *rad52* codificano per due funzioni richieste, rispettivamente, per il pathway del *Post-Replication Repair* (PRR) e della riparazione tramite ricombinazione (HR) (*vedi* testo). Il PRR è richiesto per la risposta a radiazioni UV durante la replicazione del DNA e pertanto mutanti *rad6* sono anch'essi sensibili alle radiazioni UV (B). Al contrario, la mutazione *rad52* da sola non causa sensibilità al trattamento con raggi UV (C). Se si combinano sia la mutazione *rad6* che la mutazione *rad52* con la mutazione *rad25*, i doppi mutanti (B e C) sono più sensibili dei singoli mutanti perché le mutazioni appartengono, come già detto, a gruppi di epistasi diversi.

dei geni codificanti per proteine richieste nella riparazione dei danni causati da raggi UV negli eucarioti sono stati identificati utilizzando il lievito *Saccharomyces cerevisiae* e isolando mutanti sensibili al trattamento con raggi UV a dosi in cui la sopravvivenza dell'organismo wild-type non è significativamente alterata (fig. 7.6). La combinazione di diverse mutazioni e l'analisi genetica della sensibilità a un certo agente mutageno (ad es. il trattamento con raggi UV) permettono di raggruppare le diverse mutazioni in **gruppi di epistasi**. Due geni vengono normalmente indicati come appartenenti allo stesso gruppo di epistasi quando un doppio mutante nei due geni mostra una sensibilità (ad es. ai raggi UV) simile a quella del singolo mutante più sensibile (fig. 7.6A). Diversamente, se la combinazione delle due mutazioni dà un effetto additivo, si considera che i due geni appartengano a gruppi di epistasi distinti. L'interpretazione più semplice di queste osservazioni presuppone che geni appartenenti a gruppi di epistasi distinti siano coinvolti in meccanismi molecolari alternativi per riparare danni causati dallo stesso agente mutageno; al contrario, geni appartenenti allo stesso gruppo di epistasi potrebbero codificare per singoli polipeptidi facenti parte di complessi multiproteici o per fattori proteici che agiscono sequenzialmente nello stesso pathway biochimico.

I geni identificati in lievito sono conservati anche in altri eucarioti, incluso l'uomo. Mutazioni in geni umani codificanti per proteine del NER sono associate ad almeno tre malattie genetiche, note come *Xeroderma pigmentosum* (XP), *sindrome di Cockayne* (CS) e *tricotiodistrofia* (TTD). Oltre a una serie di sintomi particolari, tutte e tre queste malattie hanno in comune un'estrema sensibilità dei pazienti alla luce del sole.

Il meccanismo del NER si suddivide in due pathway che sono, nella prima parte, distinti uno dall'altro, per poi convergere in un meccanismo comune (fig. 7.7). Il **Global Genome NER** (GG-NER) controlla l'intero genoma alla ricerca di eventi (quali i dimeri di pirimidina) che causano una distorsione della doppia elica; il secondo pathway è chiamato **Transcription-Coupled Repair NER** (TCR-NER) in quanto agisce su danni localizzati nelle regioni trascritte dalle

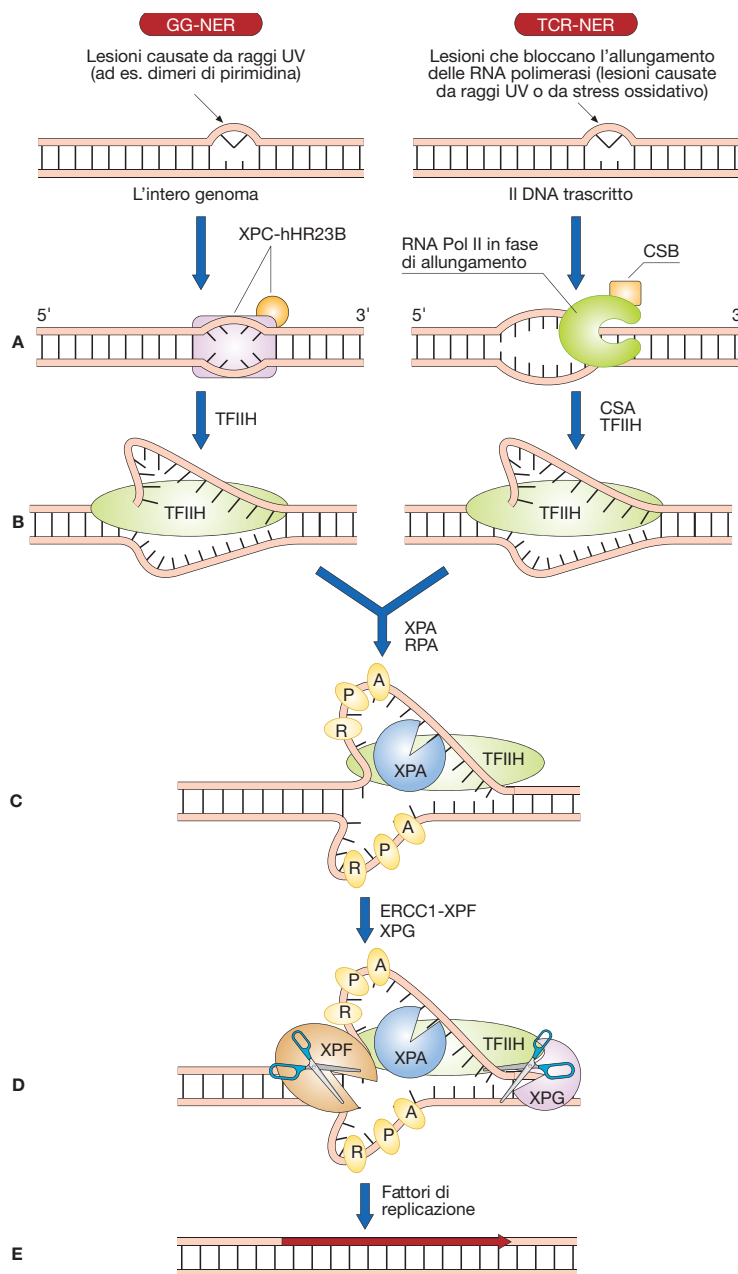


Figura 7.7 Meccanismo del NER nei mammiferi. In tutti gli eucarioti il NER è diviso in due sotto-pathway chiamati *Global Genome NER* (GG-NER) e *Transcription-Coupled Repair* (TCR-NER). I due sotto-pathway si distinguono nelle fasi iniziali del processo per poi convergere in un unico meccanismo comune. Nel GG-NER (fase A, a sinistra), il complesso proteico XPC-hHR23B più che ricercare specifiche lesioni sul DNA, individua regioni in cui il corretto appaiamento tra i due filamenti del DNA è alterato a causa di una distorsione della doppia elica. Nel TCR-NER (fase A, a destra) diversi tipi di lesione nelle regioni trascritte determinano un blocco delle RNA polimerasi (particolarmente studiata è l'RNA polimerasi II) che devono essere rimosse dal DNA per permettere la riparazione dei danni. Questo evento richiede due proteine specifiche del TCR-NER, chiamate CSA e CSB. I passaggi successivi del GG-NER e del TCR-NER sono essenzialmente identici. Il complesso multiproteico TFIIH, che contiene due polipeptidi chiamati XPB e XPD con attività DNA elicastica (fase B), apre il DNA per un tratto di circa 30 nt (nucleotidi) attorno alla lesione. La proteina XPA conferma

la presenza della lesione legandosi a essa (fase C), altrimenti tutta la reazione termina a questo stadio. La regione del DNA aperta dall'azione dell'elicasa è stabilizzata dal legame con la proteina RPA, che è un fattore coinvolto anche nella replicazione del DNA e che ha un'elevata affinità di legame per il DNA a singolo filamento. XPG e ERCC1/XPF sono due endonucleasi NER-specifiche che tagliano, rispettivamente, in 3' (destra) e in 5' (sinistra) il filamento di DNA contenente la lesione, generando un frammento di circa 30 nt che verrà poi eliminato (fase D). È stato così generato un buco di 30 nt con un'estremità 3'OH nel filamento danneggiato che può ora essere riparato dai normali enzimi che replicano il DNA copiando il filamento complementare che non è stato danneggiato (fase E). Studi *in vivo*, in cui si riesce a valutare la cinetica di legame dei diversi componenti del NER al DNA danneggiato, dimostrano che i singoli fattori proteici si assemblano nell'ordine indicato nella figura: la riparazione della lesione richiede, nel suo insieme, alcuni minuti, dopo di che l'intera "macchina" riparativa viene disassemblata.

RNA polimerasi. Numerose proteine sono coinvolte in questo processo e il loro nome deriva dal fatto che sono codificate da geni che, se mutati, causano forme diverse di *Xeroderma pigmentosum* (XPA, XPC ecc.) o di *sindrome di Cockayne* (CSA e CSB). Alcune delle proteine del NER fanno parte di un complesso multiproteico, chiamato **TFIIH**, che prende il suo nome dal fatto che è uno dei fattori di trascrizione basali più importanti richiesti per la trascrizione da parte della RNA polimerasi II (vedi par. 11.4): tale osservazione indica che le stesse proteine sono, spesso, coinvolte in più di un processo del metabolismo del DNA.

Le proteine coinvolte nel GC-NER o nel TCR-NER e il loro ruolo nella reazione di riparazione sono descritti in dettaglio nella didascalia di **figura 7.7**.

Riparazione di errori replicativi (MMR)

Durante la replicazione del DNA, l'apparato replicativo può compiere, anche se molto raramente, degli errori:

- può inserire un nucleotide che non rispetta la complementarietà delle basi prevista dalla struttura del DNA;
- può provocare delezioni/inserzioni di nucleotidi in corrispondenza di particolari sequenze sul DNA che sono ripetute più volte.

Tali errori replicativi sono riparati da un sistema noto come **MMR** (*MisMatch Repair*) (**fig. 7.8**). Abbiamo già descritto il meccanismo molecolare alla base del MMR in *E. coli* (**finestra 7.1**). Negli eucarioti il meccanismo del MMR è alquanto complesso e richiede, oltre all'azione di proteine specifiche (MutSα e β, MutLα e β), anche la funzione di proteine normalmente coinvolte nei meccanismi di replicazione del DNA. Ovviamente anche negli eucarioti è necessario discriminare il filamento di DNA neosintetizzato (che contiene l'errore e deve, quindi, essere riparato) da quello utilizzato come stampo. In *E. coli* tale segnale è conferito dallo stato di metilazione del DNA: alcune basi del DNA sono metilate, ma durante la replicazione il filamento neosintetizzato rimane non metilato per un certo periodo di tempo, permettendo così il riconoscimento del filamento di neosintesi e la sua eventuale riparazione in caso di errori replicativi (**finestra 7.1**).

Negli eucarioti la metilazione del DNA non sembra giocare un ruolo altrettanto importante nella discriminazione tra filamento parentale e filamento di neosintesi, anche perché il DNA di alcuni eucarioti non è metilato, eppure il MMR è molto efficiente. Dati recenti suggeriscono che probabili interazioni proteina-proteina tra fattori di replicazione e proteine dell'MMR possano giocare un ruolo rilevante nella scelta del filamento che deve essere riparato. Inoltre, il meccanismo di discriminazione sembra essere diverso per il filamento leading rispetto a quello lagging. È stato osservato che in oltre il 60% dei pazienti che soffrono di tumore ereditario al colon (HNPCC, *Hereditary Non-Polyposis Colon Cancer*) sono presenti mutazioni in geni che codificano per proteine dell'MMR. Il ruolo dell'MMR nel processo di tumorigenesi è anche comprovato da modelli murini: l'inattivazione di geni dell'MMR non preclude il processo di embriogenesi nel topo, ma i topini mutati sono pronti all'insorgenza di tumori.

Meccanismi di tolleranza al danno (PRR)

Cosa succede alla macchina replicativa quando incontra una lesione (ad es. un dimero di pirimidina non riparato dal NER) sul filamento da copiare se il danno in questione non è stato riparato prima della replicazione? È stato dimostrato che la maggior parte delle lesioni sul DNA rappresenta un ostacolo insuperabile per l'apparato di replicazione, che, quindi, tende a bloccarsi di fronte al danno stesso. Un blocco della replicazione sarebbe però letale per la cellula, che, di conseguenza, ha evoluto dei meccanismi di tolleranza al danno che permettono

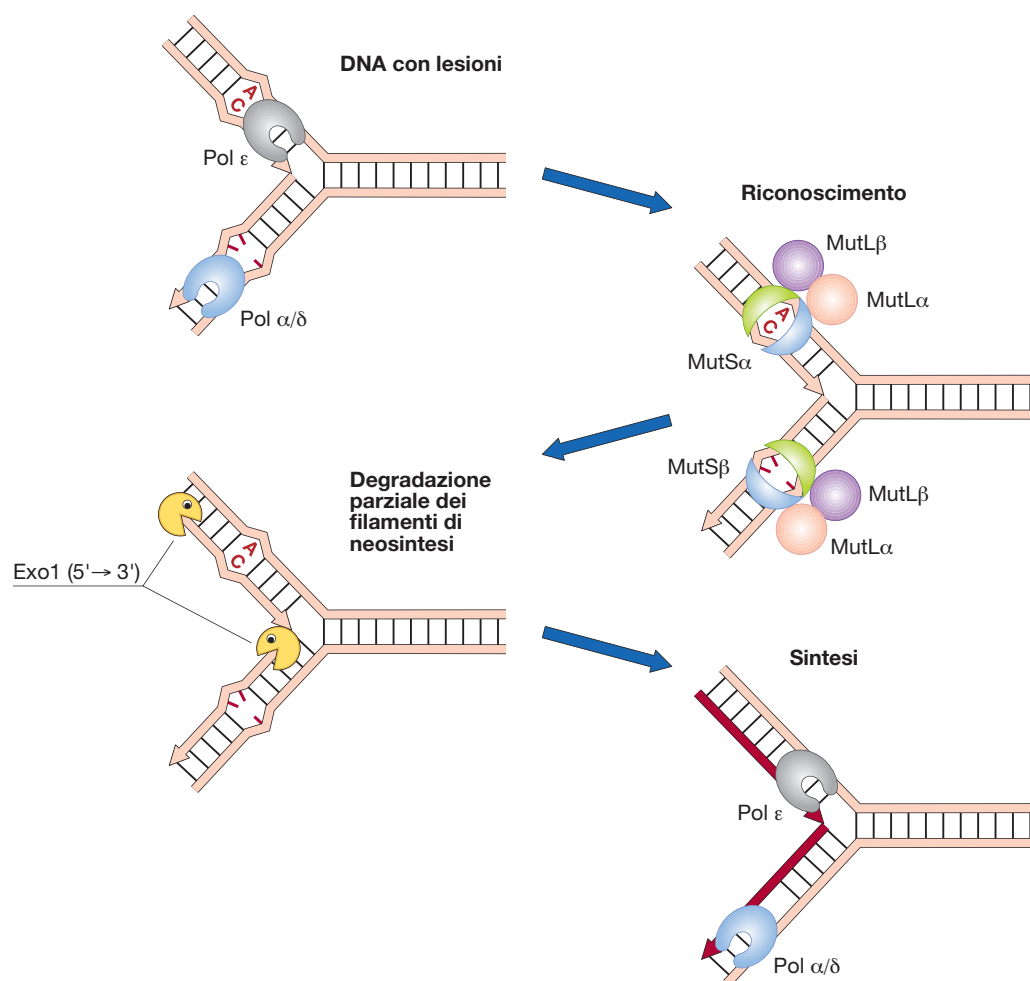


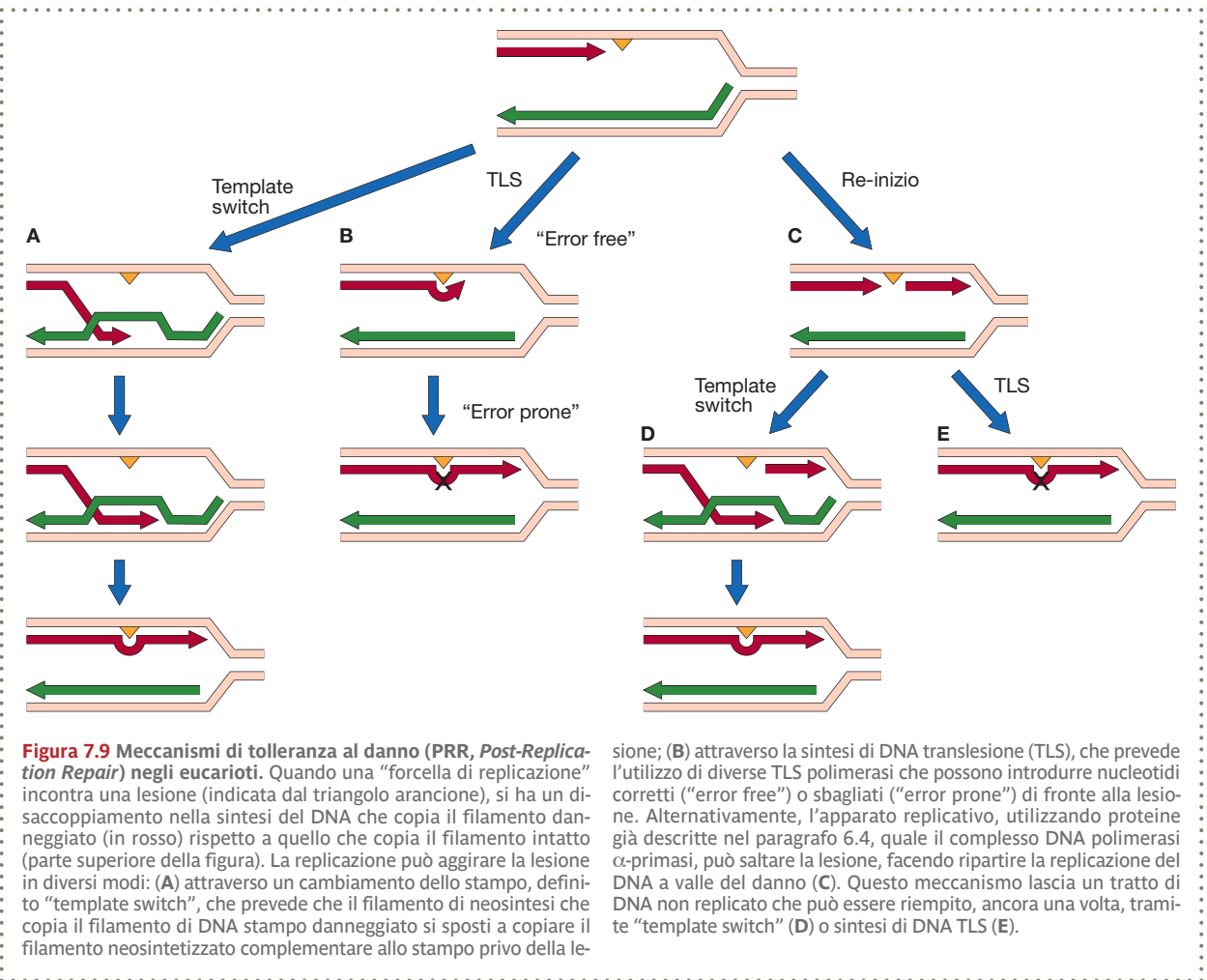
Figura 7.8 Mismatch repair (MMR) negli eucarioti. L'apparato enzimatico che agisce durante la replicazione del DNA (qui sono indicate le tre polimerasi principali, DNA polimerasi α , δ e ϵ) può compiere alcuni errori. Per esempio, sul filamento neosintetizzato può venire inserita una base sbagliata, generando l'appaiamento non complementare A-C (indicato in rosso), o si possono formare delle piccole bolle (indicate in rosso) dovute allo scivolamento dell'apparato replicativo in corrispondenza di sequenze nucleotidi-

che ripetute. Tali strutture anomale sono, dapprima, riconosciute dai complessi chiamati MutS α e MutS β che reclutano, successivamente, i complessi MutL α e MutL β . Il filamento di DNA neosintetizzato contenente gli errori replicativi viene riconosciuto e processato da una esonucleasi (Exo1) che degrada il filamento di DNA in direzione 5' → 3'. Successivamente, attraverso interazioni con tutto l'apparato di replicazione, viene ri-sintetizzato il filamento di DNA corretto (in rosso).

la replicazione di un DNA stampo danneggiato; spesso, tali meccanismi sono indicati, nel loro insieme, come **PRR** (*Post-Replication Repair*).

Le opzioni che l'apparato replicativo riesce ad attuare per aggirare una lesione sono diverse e riassunte in **figura 7.9**.

Il meccanismo noto come **template switch** è, di fatto, un meccanismo di ricombinazione (*vedi* oltre) che aggira la lesione copiando il filamento corretto presente sul cromatide fratello. La lesione originaria sul filamento di DNA danneggiato rimane, ma resta anche l'opzione di poterla riparare dopo la replicazione utilizzando un meccanismo classico, come NER, BER ecc. Potenzialmente, quindi, il danno è riparabile e non si introduce alcuna mutazione nel DNA (meccanismo "error free"). L'altra via utilizzata dalla cellula per aggirare la lesione è chiamata **sintesi di DNA translesione** (TLS). In questo caso la lesione viene aggirata utilizzando la varietà di TLS polimerasi descritte nel paragrafo 6.5. Queste polimerasi sono in grado di introdurre uno o pochi nucleotidi di fronte



sione; (B) attraverso la sintesi di DNA translesione (TLS), che prevede l'utilizzo di diverse TLS polimerasi che possono introdurre nucleotidi corretti ("error free") o sbagliati ("error prone") di fronte alla lesione. Alternativamente, l'apparato replicativo, utilizzando proteine già descritte nel paragrafo 6.4, quale il complesso DNA polimerasi α -primasi, può saltare la lesione, facendo ripartire la replicazione del DNA a valle del danno (C). Questo meccanismo lascia un tratto di DNA non replicato che può essere riempito, ancora una volta, tramite "template switch" (D) o sintesi di DNA TLS (E).

alle diverse lesioni. Esistono TLS polimerasi specializzate nell'aggirare tipi diversi di danno: alcune sono fortemente mutagene ("error prone") poiché introducono qualunque nucleotide di fronte alla lesione, mentre altre sono abbastanza fedeli e, quindi, definite come "error free".

Riparazione di rotture su entrambi i filamenti del DNA

Fra tutti i vari tipi di danno al DNA, la lesione probabilmente più pericolosa è la formazione di una rottura su entrambi i filamenti del DNA, chiamata **DSB** o **Double Strand Break**. DSB sono generati da diversi agenti chimico-fisici (in particolare radiazioni ionizzanti) e possono anche generarsi durante la replicazione del DNA se il filamento di neosintesi incontra un'interruzione ("nick") sul filamento stampo da copiare. Pochi DSB non riparati sono letali per una cellula di mammifero e un singolo DSB è mortale per una cellula eucariotica unicellulare, quale il lievito *Saccharomyces cerevisiae*.

La risposta cellulare ai DSB è estremamente complessa e collegata ai meccanismi di checkpoint che verranno descritti poco più avanti nel capitolo. Come rappresentato in **figura 7.10**, la riparazione fisica dei DSB avviene principalmente mediante un meccanismo di ricombinazione tradizionale (vedi cap. 8) o mediante giunzione diretta delle estremità rotte.

Il primo meccanismo è chiamato **HR** (*Homologous Recombination*) e coinvolge

l'azione di numerose proteine ricombinative, delle quali sono qui indicate solo le più importanti. Dagli schemi mostrati in **figura 7.10A** si può evincere che l'HR richiede l'esistenza di una copia di DNA identica a quella contenente la rottura e che la copia intatta di DNA viene usata come stampo per riparare quella contenente il DSB. La giunzione diretta delle estremità di un DSB può avvenire an-

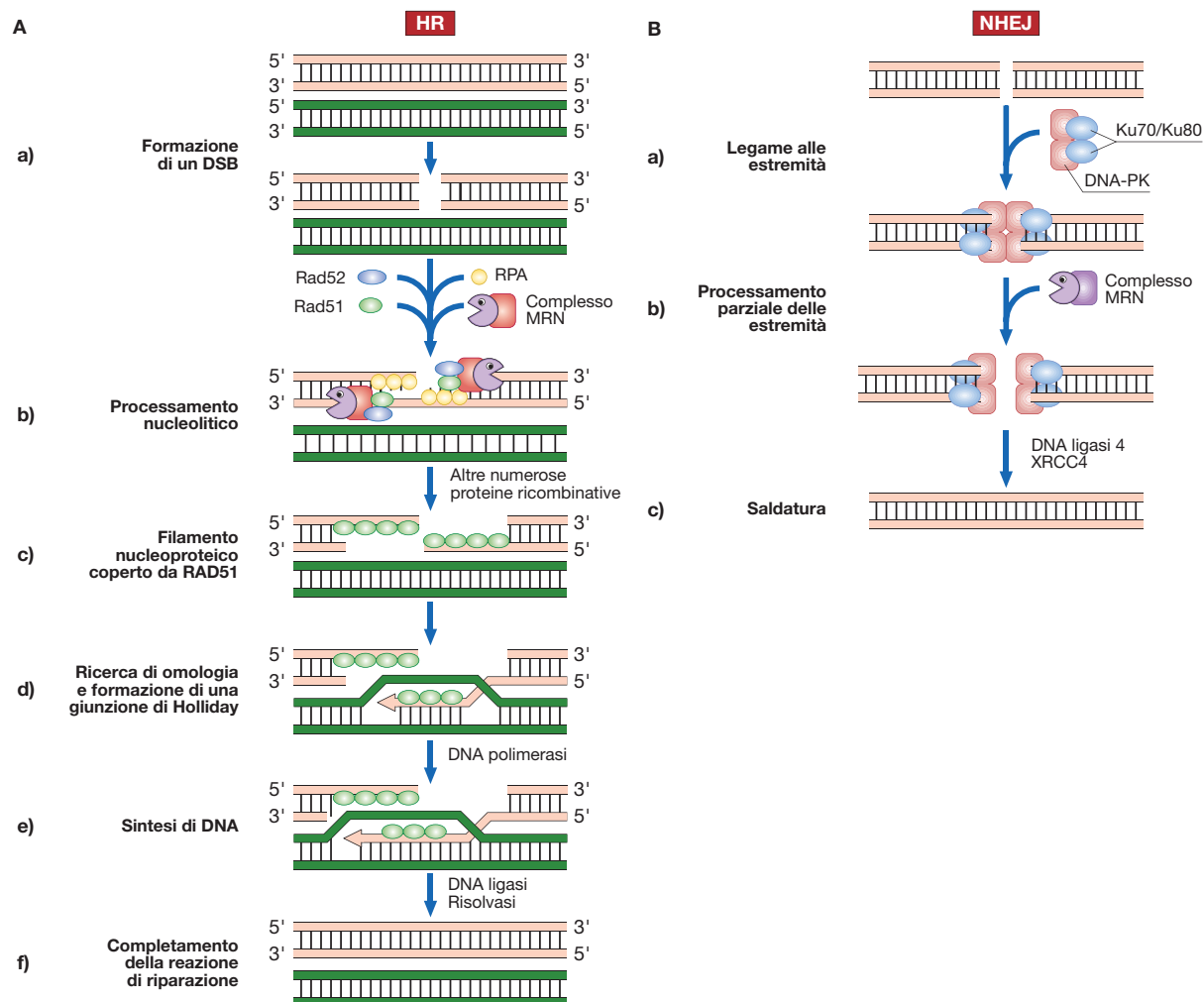


Figura 7.10 Riparazione dei DSB (*Double Strand Break* o *rotture su entrambi i filamenti del DNA*) negli eucarioti. La riparazione di DSB avviene tramite meccanismi di ricombinazione alquanto complessi. Nella figura i meccanismi sono stati semplificati e sono indicate solo le proteine più note e importanti che agiscono in tale processo nelle cellule di mammifero. I DSB sono riparati attraverso due vie principali: la prima via (A, a sinistra) è chiamata "Homologous Recombination" o HR e richiede l'azione coordinata di numerose proteine, la seconda via (B, a destra) è chiamata NHEJ (*Non-Homologous End Joining*). (A) Nel meccanismo di HR, il complesso eterotrimerico Mre11/Rad50/Nbs1 (chiamato MRN) con l'ausilio di altre proteine (in questa figura sono indicate solo Rad51 e Rad2) modifica il DSB originario causando la degradazione in direzione 5' → 3' delle due estremità 5' di ciascun DSB. Si generano, così, delle regioni di DNA a singolo filamento che sono stabilizzate dal legame della proteina RPA (fase b). Successivamente la proteina Rad51, con l'ausilio di Rad2 e altre proteine ricombinative, scalza RPA e forma un filamento nucleoproteico in cui il DNA a singolo filamento è ricoperto da Rad51 (fase c). L'estremità 3'OH del filamento coperto da Rad51, con l'aiuto di altre proteine ricombinative,

procede ora alla ricerca di sequenze omologhe di DNA presenti sul cromosoma omologo (o sul cromatide fratello) non danneggiato, formando quella che viene chiamata una giunzione di Holliday (fase d e vedi par. 8.2). La perdita di Rad51 all'estremità 3'OH innescava la neosintesi di DNA (fase e); la risoluzione degli intermedi di ricombinazione da parte di proteine specifiche, qui genericamente indicate come risolvasi, e l'azione della DNA ligasi generano due molecole riparate (fase f). La figura mostra uno dei diversi modi in cui può realizzarsi il meccanismo di ricombinazione innescato da un DSB (vedi par. 8.2). Il meccanismo qui indicato è chiamato SDSA (*Synthesis Dependent Strand Annealing*) e prevede che il DNA neosintetizzato si stacchi dopo un po' dallo stampo che sta copiando e ritorni sulla molecola di DNA originariamente danneggiata. (B) Nel processo di NHEJ le estremità del DSB sono riconosciute da un complesso noto come Ku70/Ku80 che interagisce con una proteina chinasi, chiamata DNA-PK (fase a). Può accadere che le estremità del DSB siano processate in modo limitato localmente con il possibile intervento del complesso MRN (fase b). Infine, sono reclutate le proteine XRCC4 e la DNA ligasi 4 la cui azione porta alla saldatura diretta delle estremità rotte (fase c).

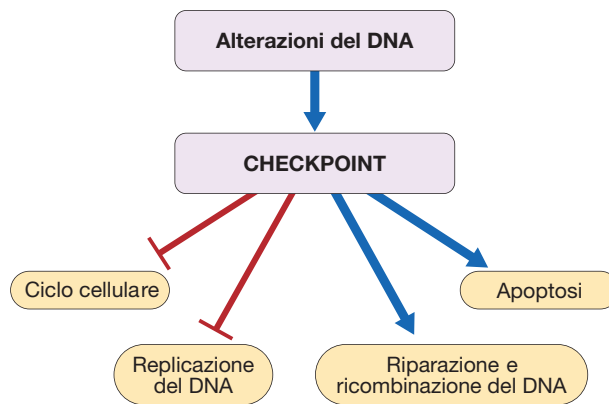


Figura 7.11 La risposta cellulare a danni al DNA attiva il checkpoint. In risposta a danni al DNA vengono attivati dei meccanismi di sorveglianza geneticamente controllati, chiamati “checkpoint”, che regolano diversi aspetti del metabolismo cellulare. Per esempio, la progressione del ciclo cellulare e la replicazione del DNA vengono inibite (freccie con estremità piatte/barrate), mentre si ha una stimolazione dei meccanismi di riparazione e di ricombinazione del DNA ed, eventualmente, l’induzione del meccanismo di morte cellulare programmata o apoptosi (freccia a punta).

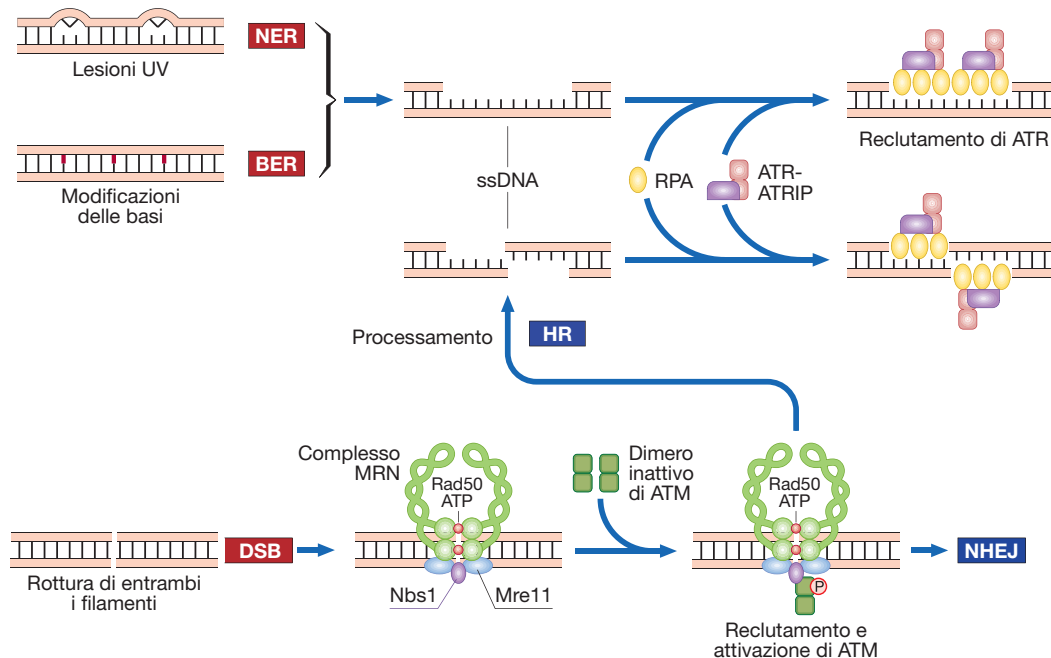


Figura 7.12 Reclutamento dei “sensori” del checkpoint da danno vicino alle lesioni. Varie lesioni sul DNA (dimeri di pirimidine, modificazioni delle basi azotate, DSB) vengono riconosciute da proteine coinvolte nei rispettivi processi di riparazione (NER, BER, HR) e modificate per generare regioni di DNA a singolo filamento (ssDNA) come indicato in figura. La generazione di ssDNA in risposta a rotture su entrambi i filamenti di DNA (DSB) tramite l’azione del complesso MRN (vedi anche fig. 7.10) è il meccanismo meglio conosciuto nei dettagli

molecolari. In cellule di mammifero, il legame di MRN ai DSB può innescare due vie riparative: la prima, attraverso il processamento dei DSB da parte del complesso MRN, genera regioni di ssDNA ricoperte dalla proteina RPA e il reclutamento del “sensore” di checkpoint noto come ATR-ATRIP. In alternativa, se il processamento da parte di MRN è limitato, viene reclutato il “sensore” di checkpoint noto come ATM. Tale reclutamento comporta la conversione di un dimero inattivo di ATM nella sua forma attiva costituita da un monomero fosforilato.

che tramite un meccanismo alternativo noto come **NHEJ** (*Non-Homologous End Joining*). Tale reazione (**fig. 7.10B**) è apparentemente più semplice e richiede un numero più limitato di fattori proteici. La possibilità di riparare un DSB tramite l’uno o l’altro dei due processi sopra citati è fortemente dipendente dalla fase del ciclo cellulare (vedi oltre) in cui il DSB deve essere riparato e dallo stato aploide o diploide della cellula. Infatti, è chiaro che il meccanismo di riparazione tramite HR richiede una copia di DNA intatta presente o nel cromatide fratello o sul cromosoma omologo. Mutazioni nei geni coinvolti nella riparazione dei DSB causano predisposizione a una varietà di tumori e, spesso, riarrangiamenti cro-

mosomici vistosi; la perdita della funzione di tali geni nei modelli murini causa quasi sempre letalità durante lo sviluppo embrionale a conferma dell'estrema importanza della riparazione dei DSB.

7.3 Risposta cellulare a danni sul DNA e connessioni tra riparazione e ciclo cellulare

Negli eucarioti la risposta cellulare a danni sul DNA è molto complessa e si integra con i meccanismi che controllano la progressione del ciclo cellulare, i cui aspetti più rilevanti sono stati descritti nel paragrafo 6.7. Come abbiamo visto, la corretta successione delle diverse fasi del ciclo cellulare è assicurata da meccanismi di sorveglianza, definiti checkpoint e, tra questi, uno dei più importanti è il **checkpoint da danno al DNA** o **DNA damage checkpoint**.

La conseguenza più evidente dell'attivazione del checkpoint da danno al DNA consiste in un rallentamento o in un blocco temporaneo del ciclo cellulare. In realtà, la risposta della cellula è più complessa (fig. 7.11) e non riguarda solo la progressione del ciclo cellulare, ma anche una rete di interconnessioni tra diversi aspetti del metabolismo del DNA e l'eventuale attivazione di un meccanismo di sorveglianza estremo. Se, infatti, la quantità di danni al DNA è tale da essere considerata non più riparabile, la cellula attiva un meccanismo noto come **morte cellulare programmata** o **apoptosi**: ne consegue che la cellula "si suicida" piuttosto che correre il rischio di trasmettere alle generazioni successive un'informazione genetica aberrante e, quindi, dannosa. In estrema sintesi, l'attivazione del

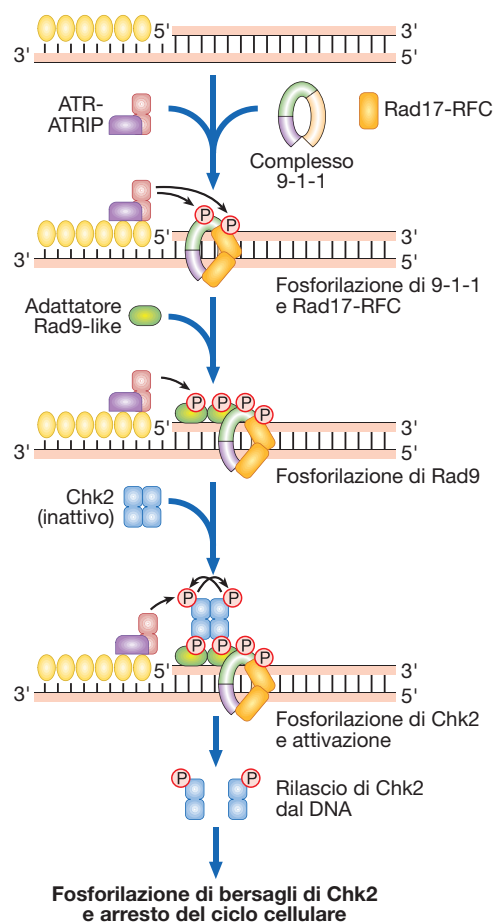


Figura 7.13 Il checkpoint da danno al DNA è una cascata di trasduzione del segnale. Il primo evento della cascata è il reclutamento di una proteina chinasi sensore (in figura è indicato il complesso ATR-ATRIP). Più o meno contemporaneamente, nelle vicinanze della lesione viene caricato il complesso noto come 9-1-1, che ha una struttura ad anello simile al PCNA e che, per essere caricato sul DNA, necessita di un "caricatore", il complesso Rad17-RFC. Subunità dei complessi 9-1-1 e Rad17-RFC sono fosforilate da ATR e tale modificazione serve per reclutare proteine che fungono da "adattatori". Il più noto adattatore è la proteina Rad9 di lievito: in mammifero vi sono più adattatori genericamente indicati in figura come Rad9-like. Rad9-like viene anch'essa fosforilata da ATR e forma degli oligomeri che determinano il reclutamento delle chinasi definite "trasmettitori del segnale". La più nota in cellule di mammifero è Chk2, che, una volta reclutata, forma degli oligomeri fosforilati da ATR. Questa modificazione attiva Chk2, che si autofosforila. Questa ulteriore modificazione coincide con la piena attivazione di Chk2, che si stacca dal DNA e va a fosforilare una serie di bersagli che agiscono come "effettori" della cascata di trasduzione del segnale bloccando, per esempio, la progressione del ciclo cellulare.

checkpoint da danno al DNA, determinando un blocco parziale o un rallentamento del ciclo cellulare, fornisce più tempo alla cellula per riparare i danni o, nel caso in cui ciò risulti impossibile, per innescare l'apoptosi.

L'interpretazione più aggiornata della risposta cellulare a danni al DNA prevede che le diverse lesioni sul DNA siano riconosciute dagli specifici sistemi di riparazione descritti precedentemente e la maggior parte dei danni venga rapidamente riparata. Tuttavia, attraverso interazioni fisiche con proteine riparative o riconoscendo particolari strutture sul DNA danneggiato, alcune proteine richieste per l'attivazione del checkpoint vengono reclutate in prossimità delle lesioni (**fig. 7.12**). Queste proteine che agiscono nella fase iniziale dell'attivazione del checkpoint da danno sul DNA vengono spesso definite come **sensori** del danno.

Come indicato in **figura 7.12**, la struttura più importante generata dal processamento di varie lesioni è la produzione di tratti di DNA a singolo filamento (ssDNA). L'ssDNA viene ricoperto da RPA (una proteina che ormai abbiamo incontrato più volte in diversi processi sul DNA) e, in prossimità di tali regioni, vengono reclutati i sensori del checkpoint. Le proteine che, nei mammiferi, agiscono agli stadi iniziali del checkpoint da danno sono chiamate **ATM** e **ATR**, e omologhi funzionali e strutturali di tali fattori proteici sono presenti in tutti gli eucarioti. ATM e ATR sono due proteine chinasi che attraverso specifiche interazioni proteina-proteina ed eventi di fosforilazione innescano la cascata di trasduzione del segnale del checkpoint (**fig. 7.13**). Infatti, le proteine "sensori" sentono la struttura generata in prossimità delle lesioni, legandosi a essa e reclutando altre proteine che agiscono essenzialmente come "adattatori" o piattaforme di reclutamento per altri polipeptidi, tra cui uno o più "trasmettitori": questi fat-

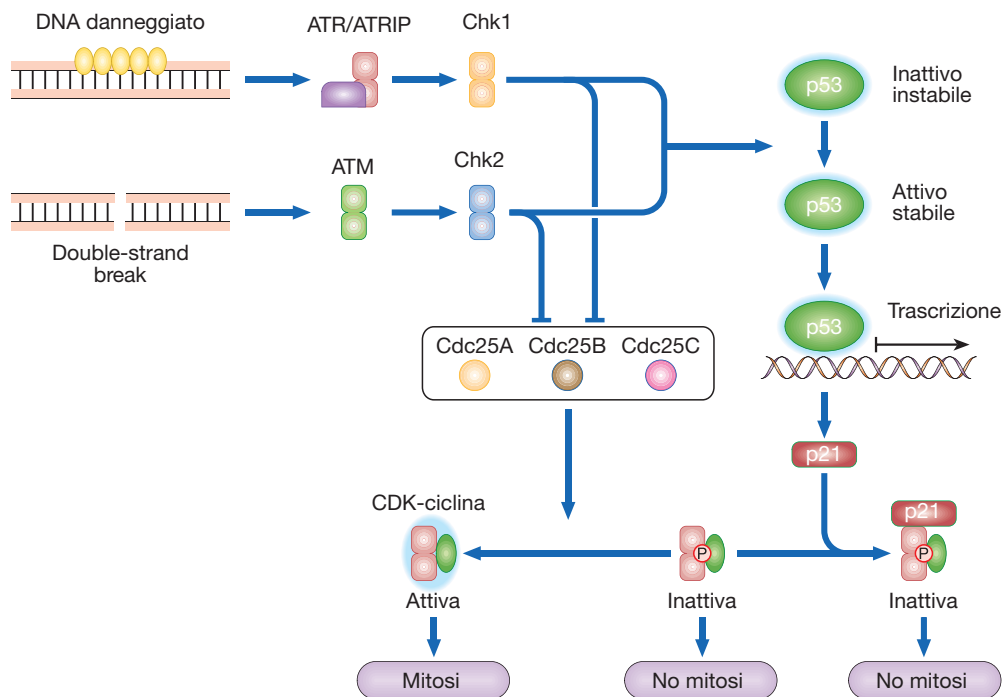
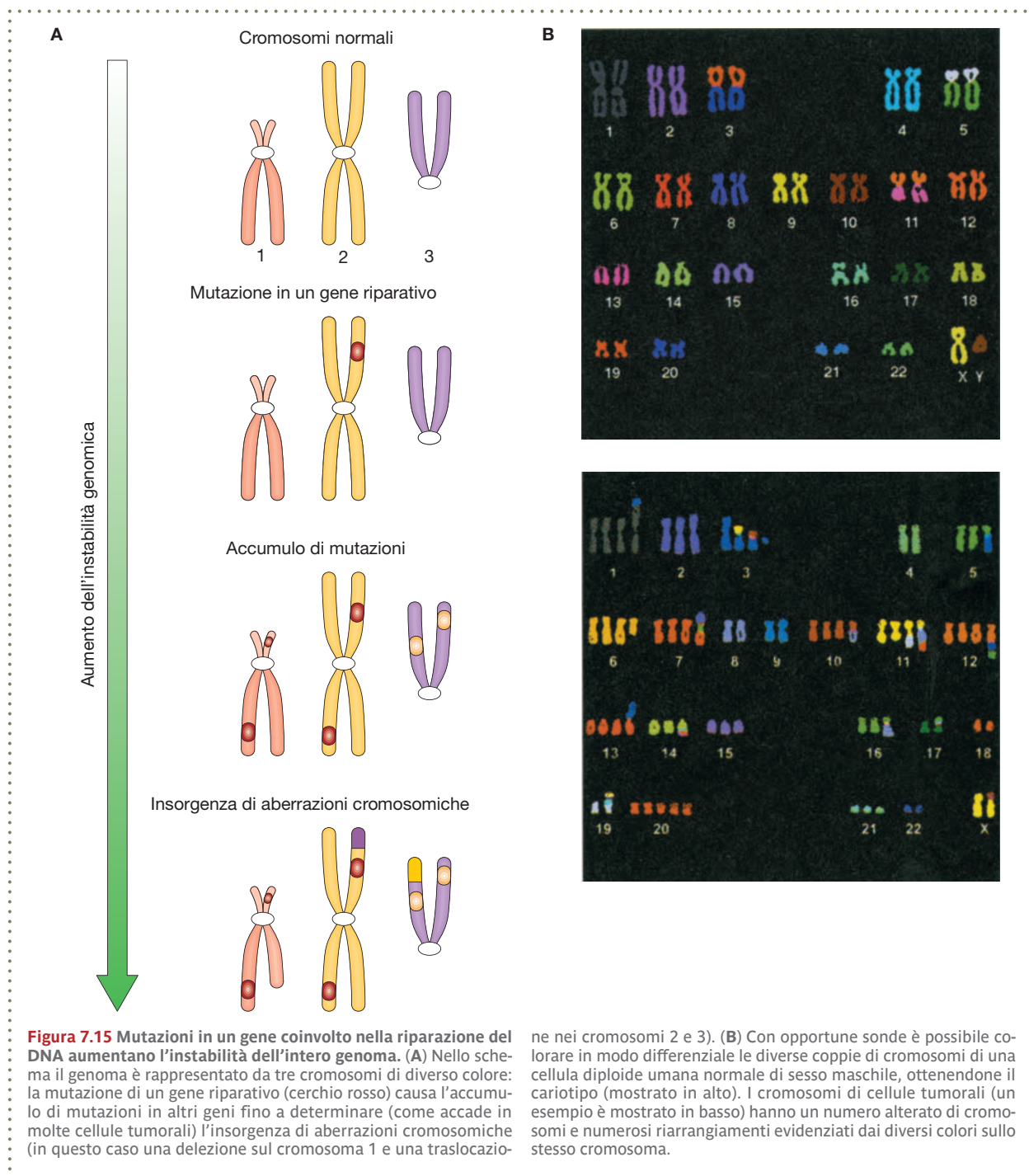


Figura 7.14 Il checkpoint da danno al DNA agisce su p53 e Cdc25 bloccando l'entrata in mitosi. L'attivazione di ATM e ATR in risposta a danni al DNA porta all'attivazione delle proteine chinasi Chk1 e Chk2, che agiscono inibendo in modo complesso i tre membri della famiglia di proteine-fosfatasi Cdc25. La defosforilazione di un residuo critico sulla CDK mitotica è richiesta per l'entrata in mitosi (vedi par. 6.7): ne deriva che, bloccando la fosfatasi,

si impedisce la mitosi stessa. L'attività della CDK mitotica viene anche inibita tramite una via più lenta dalla proteina p53. p53 viene stabilizzata e attivata da ATR e ATM: tra i geni regolati da p53 c'è il gene codificante per la CKI chiamata p21 che, legandosi alla CDK mitotica, ne impedisce la funzione. L'alone azzurro attorno a p53 e alla CDK-ciclina evidenzia quando le proteine in oggetto sono attive.

tori proteici sono anch'essi delle proteine chinasi che amplificano e trasmettono il segnale così generatosi a degli "effettori" la cui funzione è richiesta, per esempio, per controllare il passaggio da una fase all'altra nel ciclo cellulare, quali le proteine p53 e Cdc25 (**fig. 7.14**). Il coinvolgimento di queste due proteine cruciali per la progressione del ciclo cellulare ci fornisce un'idea, seppur superficiale, dell'importanza del checkpoint da danno al DNA. p53 è un fattore trascrizionale che lega il promotore di numerosi geni influenzandone il livello d'espressione. La fosforilazione ATM e ATR-dipendente di p53 stabilizza la proteina determinando, fra l'altro, un aumento della trascrizione del gene codificante p21, uno degli



inibitori delle CDK. Cdc25 è una proteina-fosfatasi che modulando lo stato di fosforilazione delle CDK ne influenza l'attività (*vedi sopra*). Ne consegue che il checkpoint modula lo “stato del motore” del ciclo cellulare agendo anche direttamente sull'attività delle CDK.

Per avere un'idea dell'importanza di tali controlli basti pensare che mutazioni nel gene che codifica per p53 sono state trovate in più del 50% di tutti i tumori dell'uomo; il concetto che un alterato controllo della stabilità del genoma è alla base della cancerogenesi è diventato uno dei paradigmi centrali della Biologia moderna e riceve continue conferme sperimentali. Per cercare di esemplificare il problema, basti pensare che il mal funzionamento di un gene che controlla una funzione essenziale per l'integrità del DNA, per esempio un gene richiesto per l'attivazione del checkpoint o la riparazione del DNA, determina un fenotipo “mutatore”. Ciò significa che un difetto nella risposta cellulare a danni al DNA causa l'accumulo di mutazioni in altri geni (**fig. 7.15**). Se questi geni regolano, a loro volta, funzioni importanti per il controllo della proliferazione cellulare, il risultato potrebbe essere una cellula che si divide in modo incontrollato. In altri termini, un'elevata instabilità genomica aumenta la velocità con cui si accumulano mutazioni: la successiva selezione di cellule con una più rapida proliferazione cellulare, e magari un'aumentata capacità di invadere i tessuti circostanti, può portare alla formazione di cellule tumorali dotate di proprietà metastatizzanti.

Lecture di approfondimento

- Bartek, J., Bartkova, J. e Lukas, J. (2007), «DNA Damage signalling guards against activated oncogenes and tumour progression», *Oncogene*, **26**, pp. 7773-7779.
- Friedberg, E.C., Elledge, S.J., Lehmann, A.R., Lindahl, T. e Muzi-Falconi, M. (Eds), *DNA repair, mutagenesis and other responses to DNA damage*, Cold Spring Harbor Laboratory Press, 2014.
- Hanawalt, P.C. (2007), «Paradigms for the Three Rs: DNA replication, Recombination and Repair», *Molecular Cell*, **28**, pp. 702-707.
- Harper, J.W. e Elledge, S.J. (2007), «The DNA damage response: ten years after», *Molecular Cell*, **28**, pp. 739-745.

8

Ricombinazione

8.1 Introduzione e ricombinazione in meiosi

8.2 Ricombinazione omologa o generalizzata

8.3 Trasposizione

8.4 Riarrangiamenti di sequenze di DNA e controllo dell'espressione genica

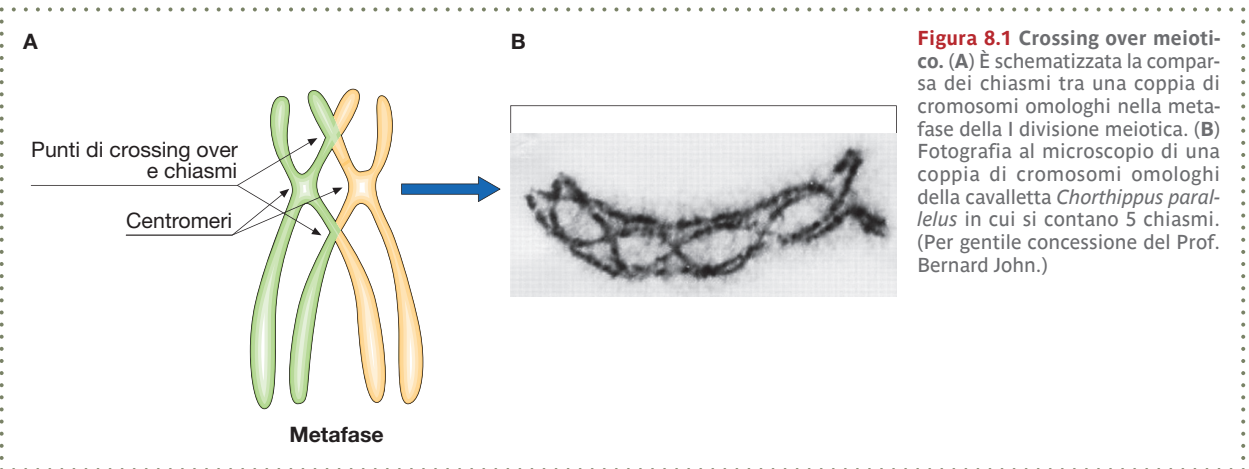
8.1 Introduzione e ricombinazione in meiosi

Tutti i genomi evolvono acquisendo nuove sequenze di DNA o riarrangiando quelle esistenti. Con il termine **ricombinazione genetica** si comprendono tutti quei processi che mediante rotture e unioni di molecole di DNA ne modificano l'organizzazione strutturale.

Abbiamo visto nel paragrafo 7.2 che attraverso meccanismi ricombinativi si possono riparare molecole di DNA che contengono una rottura a doppio filamento. Tuttavia, la funzione principale della ricombinazione è quella di creare nuove combinazioni di geni sul DNA, determinando la variabilità genetica che consente alle specie di evolversi con il cambiare delle condizioni ambientali. Questa funzione della ricombinazione è particolarmente importante in meiosi (si rimanda a un testo di Genetica) quando i cromosomi omologhi si appaiano durante la prima divisione meiotica. Come risultato di questo appaiamento, possono avvenire degli scambi fisici di tratti di DNA tra la coppia di cromosomi omologhi: il meccanismo è normalmente chiamato **crossing over** ed è osservabile anche al microscopio, dove i punti d'incrocio tra i 2 cromosomi sono chiamati **chiasmi** (fig. 8.1). La frequenza di crossing over tra due geni sullo stesso cromosoma dipende dalla distanza fisica tra le due sequenze di DNA: maggiore è la loro distanza, maggiore è la frequenza di un possibile crossing over tra i due geni. Le prime mappe genetiche che posizionano i geni sui cromosomi sono state ottenute, appunto, calcolando la frequenza di crossing over tra i vari geni: si è così stabilito che i geni sono disposti sui cromosomi secondo un ordine fisso e lineare.

Il crossing over tra una coppia di cromosomi in meiosi è l'esempio più classico del meccanismo di **ricombinazione omologa o generalizzata**: tale termine indica lo scambio di materiale genetico tra due molecole di DNA (o regioni di esse) che possiedono ampi tratti con elevata similarità di sequenza nucleotidica. Esistono altri due tipi principali di ricombinazione: la **ricombinazione sito-specifica**, in cui gli scambi si verificano solo in corrispondenza di una specifica sequenza di DNA, e la **trasposizione del DNA**, che riguarda, solitamente, un breve tratto di DNA con la capacità di spostarsi da una posizione su un cromosoma a un'altra.

I meccanismi di ricombinazione sopra citati sono molto complessi e si realizzano attraverso una varietà di vie ricombinative in cui agiscono numerosissime proteine: in questo capitolo prenderemo in considerazione solo alcuni dei mec-



canismi ricombinativi, rimandando a testi specialistici per una descrizione completa dei vari processi di ricombinazione. Riteniamo, infatti, che una descrizione seppur semplificata della ricombinazione debba essere inclusa anche in un testo di Biologia Molecolare. Infatti, la comprensione dei meccanismi ricombinativi ha permesso, da una parte, di comprendere le basi molecolari di una serie di processi biologici fondamentali e, dall'altra, di sviluppare una serie di tecnologie che consentono di modificare il materiale genetico.

Come vedremo più in dettaglio nel paragrafo 20.4, con il termine di “tecnologie del DNA ricombinante” si comprendono numerose e svariate tecniche che permettono di clonare geni, modificarli e, magari, trasferirli da un organismo all'altro. Alcune di tali tecniche hanno acquisito un'importanza straordinaria per comprendere le funzioni di tutti i geni di un organismo, e lo sviluppo delle tecnologie del DNA ricombinante, nel suo insieme, ha rivoluzionato l'approccio sperimentale dei ricercatori per cercare di comprendere fenomeni biologici complessi a livello molecolare.

8.2 Ricombinazione omologa o generalizzata

Il meccanismo di ricombinazione omologa è il più importante tra i vari processi ricombinativi e tutti gli organismi viventi, procarioti o eucarioti, condividono essenzialmente alcuni passaggi chiave visualizzati nelle **figure 8.2 e 8.3**.

1. Allineamento di due molecole di DNA omologhe che devono avere sequenze identiche o quasi per una regione di almeno un centinaio di nucleotidi.
2. Introduzione sul DNA di rotture a singolo o a doppio filamento che generano i due modelli di ricombinazione presentati, rispettivamente, nelle **figure 8.2 e 8.3**.
3. Formazione di due molecole di DNA congiunte che si formano quando un filamento con l'estremità 3'OH libera di una delle molecole di DNA trova una sequenza complementare sulla molecola di DNA omologa. Le due molecole sono ora congiunte da filamenti che si incrociano formando nella regione di omologia la struttura che viene chiamata **giunzione di Holliday**. In **figura 8.2** è schematizzata la versione del modello originario proposto da Robin Holliday che rappresenta ancora bene alcuni degli eventi centrali della ricombinazione (invasione del filamento, migrazione del chiasma, risoluzione della giunzione ecc.). Tuttavia, soprattutto come conseguenza dei risultati ottenuti sviluppando le tecnologie del DNA ricombinante, e tramite la dimostrazione che nella maggior parte degli eventi di ricombinazione c'è neosintesi di DNA, il modello oggi più accreditato di ricombinazione omologa indica che la maggior parte (se non tutti) degli eventi di ricombinazione è innescata

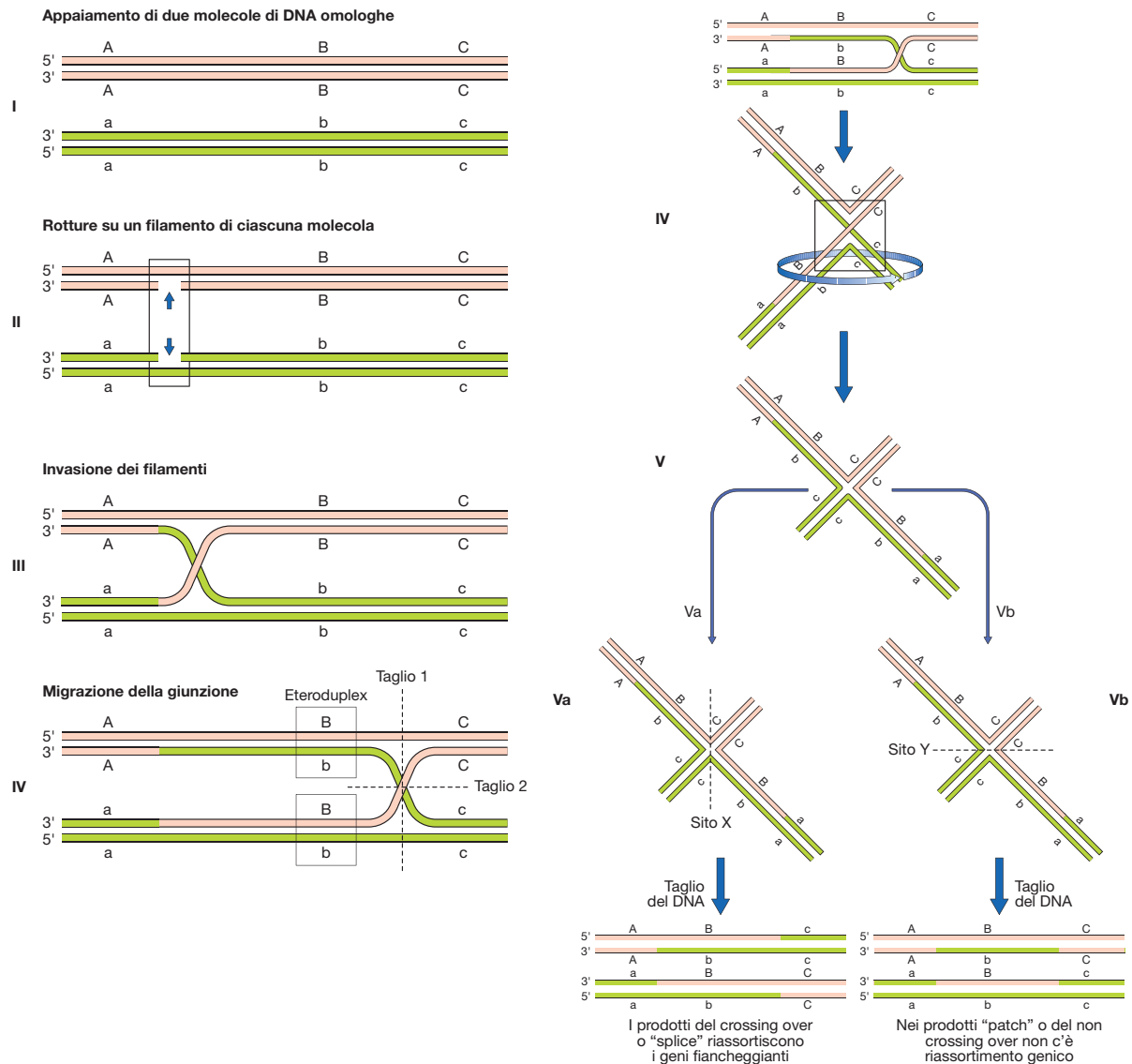


Figura 8.2 Modello originario di ricombinazione di Holliday innescato da tagli su un singolo filamento di DNA. Le due molecole di DNA omologhe sono indicate con colori diversi e le lettere maiuscole o minuscole (A, a; B, b; C, c) indicano forme alternative (alleli) di 3 ipotetici geni A, B e C. Per esempio, le sequenze di DNA dei 2 alleli A e a del gene A hanno un elevato livello di similarità (differiscono di qualche nucleotide) ma possono dare origine a fenotipi diversi. Le due molecole omologhe si appaiano (I) e la ricombinazione inizia con l'introduzione di rotture a singolo filamento (II). I filamenti incisi possono staccarsi dal filamento complementare della stessa molecola di DNA e invadere la doppia elica di DNA omologa appaiandosi al filamento complementare. Nella figura l'invasione è simmetrica e porta alla formazione di una giunzione di Holliday o chiasma (III). La giunzione di Holliday può ora scorrere sul DNA mettendo di fronte nei due filamenti di DNA sequenze non identiche (i due alleli B e b), formando delle regioni chiamate "eteroduplex" indicate dai due quadrati (IV). Per terminare l'evento di ricombinazione, la giunzione deve essere risolta

tagliando e risalando i filamenti di DNA prodotti. Per visualizzare meglio i siti di taglio la giunzione viene ruotata dando origine a una struttura planare in cui i filamenti non sono più intersecati tra loro (V). Il sito di taglio indicato con X coinvolge i due filamenti di DNA che all'inizio dell'evento ricombinativo (II) non erano rotti. Al contrario, il sito di taglio indicato con Y avviene sui due filamenti di DNA che erano stati rotti per iniziare la ricombinazione. Dopo la risoluzione della giunzione e la saldatura dei filamenti tagliati si generano prodotti diversi a seconda che il taglio sia avvenuto nella posizione X oppure nella posizione Y. Nel primo caso le molecole di DNA ricombinanti prodotte (definite "prodotti del crossing over" o "splice") hanno portato a un riassortimento degli alleli adiacenti al sito di ricombinazione (Va), perché vi è stato uno scambio tra i geni A e C. Nel caso in cui il taglio avvenga nella posizione Y, le molecole di DNA ricombinanti contengono una porzione di DNA ibrido chiamata "patch" (Vb). Tale taglio non porta al riassortimento dei geni fiancheggianti la giunzione e dà origine a prodotti definiti "prodotti del non-crossing over".

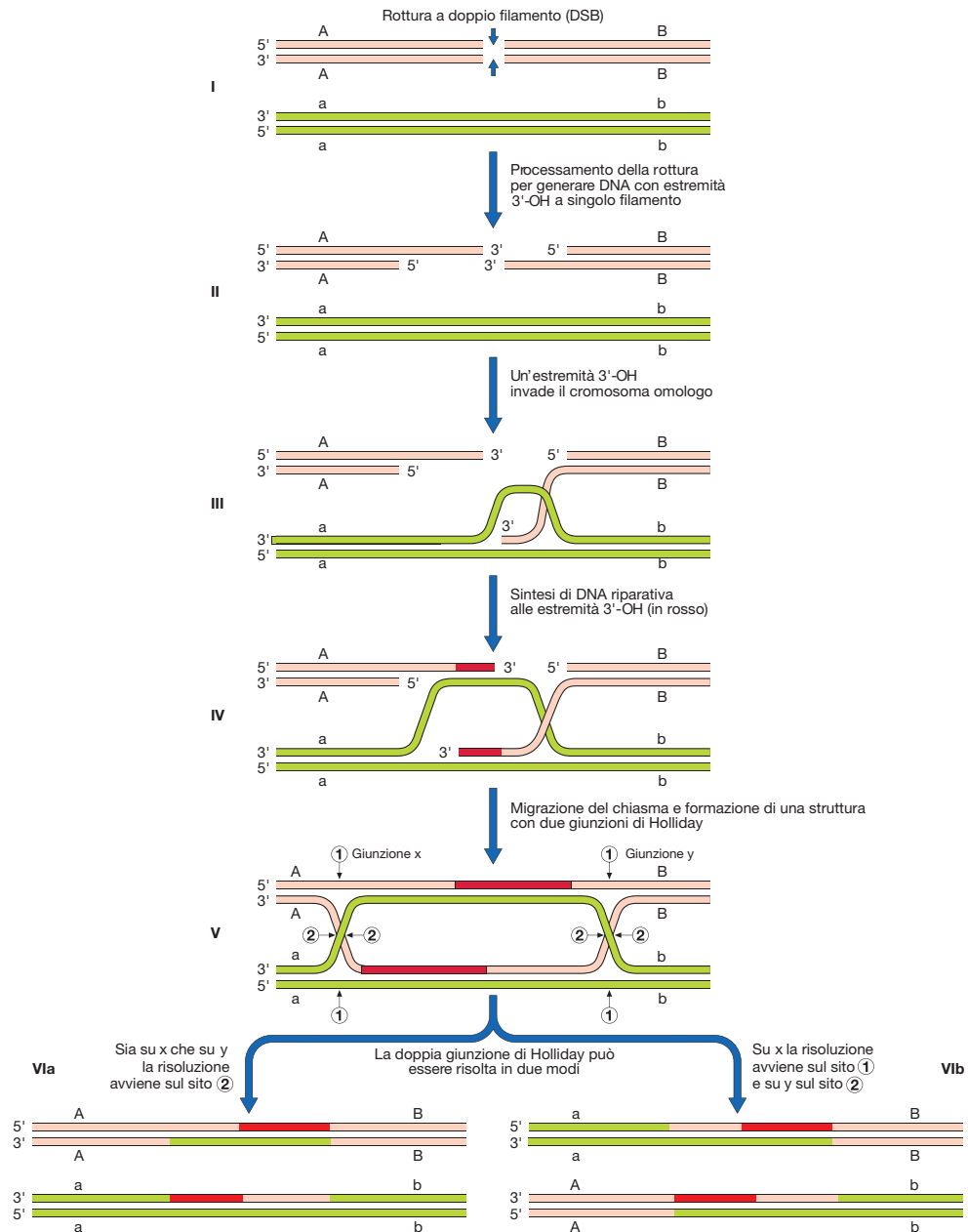


Figura 8.3 Modello di Holliday applicato a una rottura a doppio filamento (DSB) sul DNA. Nel modello di ricombinazione che coinvolge un DSB l'evento scatenante è la produzione del DSB stesso (I) (vedi anche fig. 7.10). Le estremità 5' del DSB sono processate così da generare delle estremità 3'OH sporgenti a singolo filamento (II). Una (III) o entrambe le estremità 3'OH possono invadere la molecola di DNA omologa dando origine a neosintesi riparativa del DNA (indicata in rosso) attraverso l'estensione delle estremità 3'OH e copiatura del filamento complementare intatto (IV). Se, in prossimità del sito di rottura, le sequenze di DNA delle due molecole omologhe originarie non sono identiche (alleli diversi), con il sistema di riparazione del DSB le sequenze rosa perdute durante il processamento del DSB originario possono venire rimpiazzate durante la neosintesi riparativa. Alla fine del processo di ricombinazione di un DSB può generarsi un intermedio con due giunzioni di Holliday (V), che indichiamo come x e y. Tale intermedio sarà risolto da reazioni di taglio e ricucitura come descritto in figura 8.2, e la struttura finale delle molecole ricombinanti che si formeranno è, di nuovo, influenzata dalle modalità con cui le giunzioni vengono tagliate. Per

ogni giunzione x e y ci sono due possibilità di taglio, indicate come sito di taglio 1 o 2. Se entrambe le giunzioni vengono tagliate allo stesso sito di taglio (1 o 2) si avrà la formazione di prodotti "patch" (o del non-crossing over), come mostrato nell'esempio in **VIa**, dove entrambe le giunzioni sono state tagliate al sito 2. Infatti, gli alleli A/B e a/b fiancheggiati le giunzioni si ritrovano nella stessa combinazione presente nelle molecole prima della ricombinazione (lo stesso accade se entrambe le giunzioni vengono tagliate al sito 1: infatti, anche se il taglio alla giunzione x genererebbe un prodotto di crossing over, questo effetto viene, di fatto, annullato dal taglio al sito y: crossing over + crossing over = non-crossing over). Viceversa, se le due giunzioni di Holliday sono tagliate a siti diversi (**VIb**) si ha la formazione dei prodotti di crossing over, con il riassortimento degli alleli fiancheggiati le giunzioni. Va sottolineato anche che possono esistere meccanismi alternativi a quello mostrato in figura per descrivere quanto accade a un DSB: alcuni sono più semplici (vedi l'esempio mostrato in fig. 7.10), altri, invece, più complessi. Tuttavia va al di là dello scopo di questo testo presentare tutti i meccanismi possibili.

FINESTRA 8.1

CONVERSIONE GENICA

Nel meccanismo di divisione meiotica una cellula diploide produce 4 cellule aploidi, i gameti. Quindi, una cellula diploide eterozigote per il gene *A* avrà genotipo *A/a* e, in meiosi, darà origine a 2 gameti contenenti l'allele *A* e 2 gameti contenenti l'allele *a*, con un rapporto 2:2. Nei funghi, quali *N. crassa* o *S. cerevisiae*, i prodotti di ogni singola meiosi (spore) sono contenuti in una struttura fisica detta “asco” (vedi par. 22.1). È possibile recuperare da ogni asco le singole spore e analizzare il loro genotipo. Da tali studi si è trovato che in alcuni casi c'era uno scostamento evidente dalla segregazione 2:2 attesa per un eterozigote *A/a*: accadeva che i due alleli segregavano in rapporto 3:1 o 1:3.

Come si può spiegare a livello molecolare il fenomeno della conversione genica? Riprendiamo il modello della riparazione di una rottura a doppio filamento (DSB) sul DNA mediante ricombinazione omologa. La riparazione di un DSB può realizzarsi attraverso diversi meccanismi molecolari: qui paragoniamo il meccanismo classico che dà origine a una doppia giunzione di Holliday con un altro meccanismo definito *Synthesis-Dependent Strand Annealing* o **SDSA** (fig. 8F1.1). I primi passaggi del processo (A) sono uguali nei due meccanismi: c'è un processamento delle estremità del DSB, l'invasione di un'estremità 3' a singolo filamento del cromosoma omologo per generare quello che è chiamato un D-loop. Nel modello che genera una doppia giunzione di Holliday (B), il D-loop viene esteso tramite sintesi di DNA; anche l'altra estremità viene coinvolta nel meccanismo ricombinativo e dopo sintesi di DNA e reazione di ligazione si genera un intermedio con due giunzioni di Holliday. La risoluzione casuale di tale struttura determinata da enzimi chiamati “risolvasi” genera dei prodotti definiti come non-cross over (prodotti di conversione genica) o prodotti cross over (vedi testo). La maggior parte degli eventi di conversione genica è determinata dal fatto che durante il meccanismo ricombinativo possono generarsi delle regioni eteroduplex dovute alla presenza di alleli diversi nei due omologhi coinvolti (vedi testo). Queste regioni possono venire riparate attraverso il meccanismo del mismatch repair (vedi par. 7.2 e fig. 7.8). In questo caso, però, il mismatch repair sceglie a caso il filamento da riparare. Ne deriva che un allele può venire perduto e l'altro duplicato, portando alla “conversione” netta di un allele nell'altro e spiegando i fenomeni di segregazione 3:1 o 1:3 descritti sopra.

Il meccanismo di ricombinazione omologa tramite la formazione di una doppia giunzione di Holliday predice, però, un numero quasi uguale di prodotti di cross over e di prodotti di non-cross over e non spiega la bassa frequenza (< 8%) degli eventi di cross over rispetto agli eventi di conversione genica osservati nei funghi come risultato della ricombinazione indotta da rotture di DNA a doppio filamento. Una spiegazione viene dal secondo meccanismo descritto in figura, il meccanismo SDSA (C). In tale pathway ricombinativo, dopo l'estensione del D-loop, il filamento neosintetizzato si sposta dallo stampo e si appaia con l'altra estremità 3' dello stesso omologo con successiva neosintesi di DNA e ligazione. Nel meccanismo SDSA vengono generati solo prodotti non-cross over e tutta la neosintesi del DNA avviene sullo stesso omologo. Tale meccanismo spiega l'alta frequenza nei funghi degli eventi di conversione genica.

Da un punto di vista evolutivo, la conversione genica ha verosimilmente avuto un ruolo importante nella creazione di famiglie multigeniche, e la conversione genica di un allele mutato legato a una certa patologia potrebbe portare alla reversione del fenotipo patogeno, un'ipotesi spesso definita come **terapia genica naturale**.

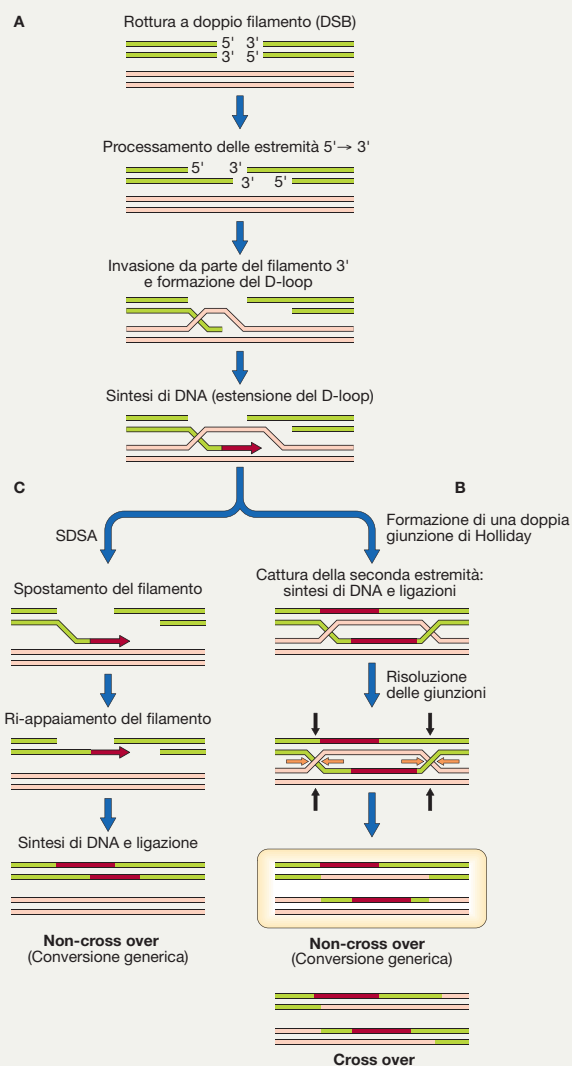


Figura 8F1.1 Eventi di conversione genica si realizzano attraverso due meccanismi di ricombinazione. Il modello della riparazione di un DSB mediante ricombinazione omologa può avvenire con due meccanismi principali. Le prime fasi dei due pathway sono comuni (A) e prevedono il processamento del DSB a generare code di DNA a singolo filamento con estremità 3' sporgenti e invasione del cromosoma omologo a formare un D-loop. Poi i meccanismi si distinguono e danno origine a un intermedio con una doppia giunzione di Holliday (B) o a un meccanismo alternativo detto SDSA (C). Dettagli sui due pathway e sui loro prodotti finali sono descritti nel testo. La risoluzione di una doppia giunzione di Holliday da parte delle resolvasi determina la produzione di un eguale numero di eventi non-cross over, cioè di conversione genica (freccie rosse a entrambe le giunzioni), e di eventi di cross over (freccie nere a una giunzione e freccie rosse all'altra giunzione).

da rotture del DNA a doppio filamento (DSB) come indicato in **figura 8.3**. In tale modello si creano due giunzioni di Holliday.

4. La/e giunzione/i di Holliday possono muoversi sulle molecole di DNA, prevenendo, nel caso del DSB mostrato in **figura 8.3**, neosintesi di DNA attraverso un processo chiamato **migrazione della giunzione** o *branch migration*.
5. La/e giunzione/i di Holliday devono, infine, essere risolte da tagli e ricuciture del DNA tramite l'azione di più proteine che qui indichiamo genericamente come **risolvasi**. La risoluzione della giunzione di Holliday può avvenire mediante tagli con orientazioni diverse che danno origine a diversi prodotti di ricombinazione (**figg. 8.2 e 8.3**).

DSB che possono innescare la ricombinazione omologa accadono frequentemente sul DNA. Abbiamo visto nel paragrafo 7.2 che una serie di agenti chimico-fisici, quali le radiazioni ionizzanti, causano tali tipi di lesioni sul DNA; un DSB può generarsi anche quando una forcella di replicazione incontra un taglio a singolo filamento sulla molecola di DNA da replicare. La cellula possiede una serie di proteine specializzate per compiere la ricombinazione omologa e alcune di queste le abbiamo indicate descrivendo il meccanismo attraverso il quale la cellula può riparare rotture a doppio filamento sul DNA (*vedi* par. 7.2).

La riparazione dei DSB può avvenire tramite meccanismi molecolari di ricombinazione omologa un po' diversi tra loro. Nella **finestra 8.1** è paragonato il meccanismo classico già presentato in **figura 8.3** con un altro meccanismo chiamato **Synthesis-Dependent Strand Annealing (SDSA)**, che offre una spiegazione molecolare a un fenomeno genetico ben conosciuto e denominato **conversione genica**.

In conclusione, la ricombinazione omologa svolge numerose funzioni nella cellula, tra cui: la riparazione di danni sul DNA, la corretta separazione e segregazione dei cromosomi in meiosi, la creazione di nuove combinazioni di geni che possono essere selezionate nel corso dell'evoluzione. Il meccanismo è così importante che nei suoi processi fondamentali è conservato in tutti gli organismi viventi: non solo negli organismi diploidi a riproduzione sessuale, ma anche nei batteri.

Ricombinazione sito-specifica

La ricombinazione conservativa sito-specifica è limitata a specifiche sequenze di DNA che possono essere anche piuttosto corte (20-200 nucleotidi). L'esempio più classico di tale tipo di ricombinazione è, probabilmente, l'integrazione del genoma del batteriofago lambda (λ) nel genoma circolare di *Escherichia coli* (**finestra 8.2**; *vedi* par. 10.4). La funzione biologica di questo specifico evento consiste nel fatto che l'integrazione del DNA del fago lambda nel genoma di *E. coli* determina l'instaurarsi di uno stato lisogenico per cui il DNA del fago viene replicato passivamente insieme al DNA batterico. Allorché il DNA del fago viene escisso dal genoma batterico si replica indipendentemente, dando origine a numerose copie che si assemblano in particelle fagiche complete che causano la lisi della cellula batterica infettata. La ricombinazione sito-specifica richiede l'azione di proteine note come **ricombinasi**, che catalizzano il taglio e la riunione delle molecole di DNA coinvolte.

Senza entrare in eccessivi dettagli, la ricombinasi coinvolta nel processo di integrazione ed escissione del DNA del batteriofago lambda è chiamata **λ integrasi** ed è codificata nel DNA del fago lambda. Sia per l'integrazione che per l'escissione del DNA del fago sono richieste altre proteine accessorie codificate sia nel DNA del fago che nel DNA del genoma di *E. coli*. I siti di ricombinazione sono chiamati **siti att**: *attP* si trova sul DNA del fago, mentre *attB* sul genoma batterico (**fig. 8.4**).

Tornando a considerare la ricombinazione sito-specifica in generale, osserviamo che essa può dare origine a tre diversi tipi di riarrangiamento sul DNA:

FINESTRA 8.2

PROPRIETÀ E CICLO VIRALE DEL FAGO LAMBDA

Il fago lambda (λ) è il rappresentante più noto di un gruppo di fagi di *Escherichia coli* definiti “temperati” (vedi par. 10.4). L’infezione di cellule batteriche da parte di tali fagi può dare origine a un ciclo litico che porta alla lisi della cellula, oppure il DNA del fago può integrarsi nel genoma del batterio (prendendo il nome di **profago**) e dare origine a un **ciclo lisogenico**. In alcune condizioni (ad es. in risposta a danni al DNA), il profago viene exciso dal genoma ospite e inizia il **ciclo litico** (fig. 8F2.1). Il capside del fago ha una forma icosaedrica e il DNA di λ nella testa del batteriofago è una molecola di DNA lineare di 48,5 kb che presenta delle estremità coesive sporgenti di 12 pb, chiamate *cos*. Quando il batteriofago infetta una cellula di *E. coli*, si attacca, tramite la coda, alla parete cellulare e inietta il suo DNA all’interno del batterio. Appena entrato all’interno dell’ospite, il DNA del fago è in grado di originare una molecola di DNA circolare a doppio filamento proprio in virtù delle sue estremità *cos*.

A questo punto il fago può dare origine a un ciclo litico oppure a un ciclo lisogenico. I dettagli molecolari che controllano tale scelta saranno descritti nel paragrafo 10.4: qui basti dire che in seguito all’infezione vengono trascritti alcuni geni di λ e la concentrazione relativa di due prodotti genici (*cro* e *CII*)

determina la scelta fra ciclo litico o lisogenico.

La replicazione del DNA di λ inizia da una specifica origine e richiede il macchinario proteico dell’ospite batterico, oltre a due proteine, chiamate O e P, codificate nel DNA virale. La replicazione tardiva del fago avviene con il meccanismo del “cerchio rotante” (*rolling circle*), per cui si generano lunghe code multimeriche contenenti più unità del genoma del fago (vedi finestra 6.5). Attraverso un processo che coinvolge numerose proteine, le code multimeriche vengono tagliate esattamente in corrispondenza delle sequenze *cos*, generando molecole di DNA lineari monomeriche. L’impacchettamento del DNA del fago all’interno del capside avviene con un meccanismo chiamato “a testa piena”: DNA delle dimensioni di circa 48,5 kb racchiuse tra due sequenze *cos* può essere impacchettato nel capside, mentre molecole di DNA più piccole o più grandi di tale dimensione non sono efficientemente maturate. Questa proprietà, unita al fatto che l’impacchettamento del DNA di λ nel capside può essere ottenuto *in vitro* da un estratto di *E. coli*, ha promosso l’utilizzo del fago lambda per la costruzione di librerie di DNA (vedi par. 20.5).

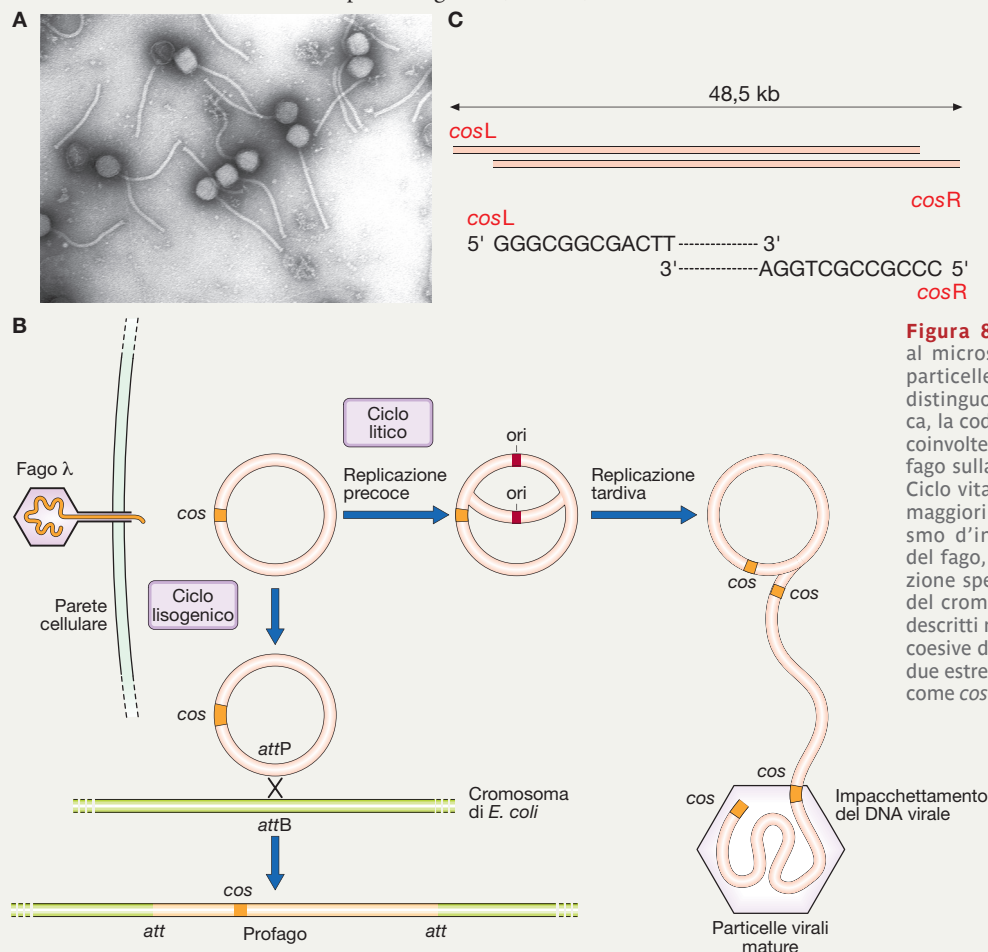


Figura 8F2.1 (A) Fotografia al microscopio elettronico di particelle del fago λ in cui si distinguono la testa icosaedrica, la coda e le fibre della coda coinvolte nell'assorbimento del fago sulla parete batterica. (B) Ciclo vitale del batteriofago λ : maggiori dettagli sul meccanismo d'integrazione del DNA del fago, mediante ricombinazione specializzata, all'interno del cromosoma di *E. coli* sono descritti nel testo. (C) Estremità coesive del DNA lineare di λ : le due estremità *cos* sono indicate come *cosR* (right) e *cosL* (left).

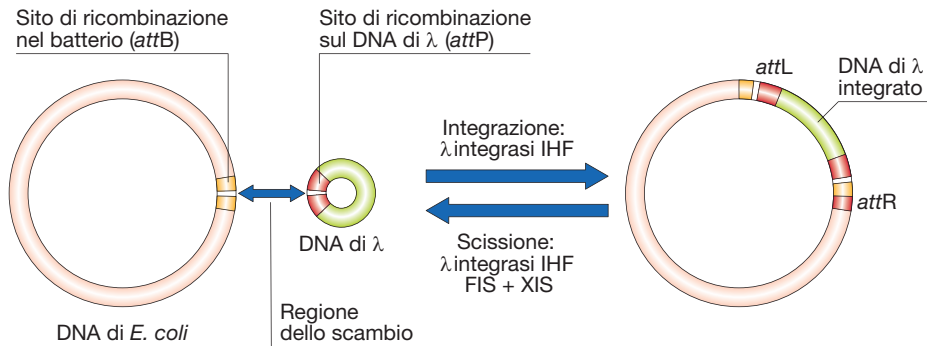


Figura 8.4 Integrazione e rimozione del DNA del batteriofago λ nel sito bersaglio sul cromosoma di *E. coli*. Il sito di ricombinazione sul DNA del fago è chiamato *attP* e il sito omologo (di soli 15 pb) sul DNA batterico è chiamato *attB*. Alla fine della reazione si generano due nuovi siti fiancheggiati il DNA del profago chiamati

attL e *attR*. La ricombinasi coinvolta è nota come λ integrasi, INT. Le proteine FIS (codificata dal batterio) e XIS (codificata dal fago) sono richieste per l'escissione del profago, insieme alla proteina IHF (codificata dal batterio).

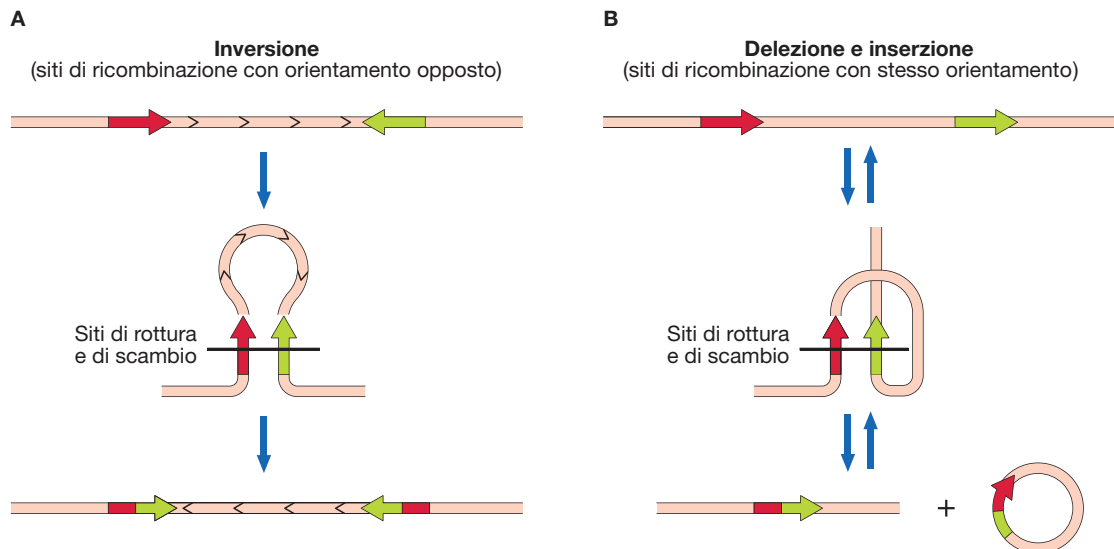


Figura 8.5 Risultati della ricombinazione sito-specifica. (A) I siti di ricombinazione (rosso e verde) hanno un orientamento opposto sulla stessa molecola di DNA: la ricombinazione tra di loro causa un'inversione. (B) Se i siti di ricombinazione (rosso e verde) sulla

stessa molecola di DNA hanno lo stesso orientamento, la ricombinazione causa una delezione; se le sequenze ricombinative con lo stesso orientamento si trovano su due molecole distinte, ne deriva un'inserzione.

1. se i siti di ricombinazione che hanno lo stesso orientamento si trovano su due molecole di DNA distinte, si ha l'integrazione di una molecola nell'altra (come nel caso dell'inserzione del DNA del fago lambda, **fig. 8.4**);
2. se i siti di ricombinazione con lo stesso orientamento si trovano su di un'unica molecola di DNA, può avvenire una delezione;
3. se i siti di ricombinazione hanno un orientamento opposto sulla stessa molecola di DNA, ne risulta un'inversione (**fig. 8.5**).

La ricombinazione conservativa sito-specifica è più comune nei batteri, dove, oltre al caso dell'integrazione del fago lambda, esistono altri esempi, quale la ricombinasi Cre codificata dal batteriofago P1, che è spesso utilizzata per l'applicazione della ricombinazione sito-specifica in tecnologie del DNA ricombinante.

Le ricombinasi coinvolte nei vari processi ricadono in due famiglie, denominate **ricombinasi in serina** e **ricombinasi in tirosina**. In entrambi i casi il mec-

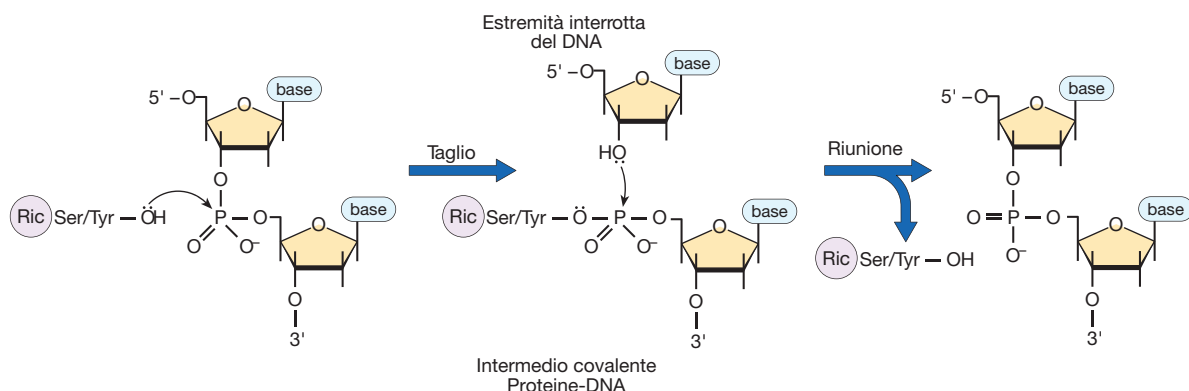


Figura 8.6 Nella loro azione le ricombinasi sito-specifiche formano un intermedio covalente e agiscono come complessi multimerici. È mostrato il gruppo OH di una serina (Ser) o tirosina (Tyr) presente nel sito attivo di una ricombinasi (Ric) mentre produce un attacco nucleofilo

sul fosfato introducendo nel sito di ricombinazione una rottura a singolo filamento e formando un intermedio covalente con il DNA. Il gruppo OH liberato sullo zucchero può attaccare il legame proteina-DNA invertendo la reazione e rilasciando la ricombinasi.

canismo d'azione è simile: il gruppo OH di una serina o di una tirosina presenti nel sito attivo della ricombinasi attacca uno specifico legame fosfodiesterico nel sito di ricombinazione. Si crea un taglio a singolo filamento sul DNA e la contemporanea formazione di un intermedio covalente tra il fosfato sul sito di taglio e il gruppo OH sulla serina o tirosina della ricombinasi (**fig. 8.6**). Questo intermedio covalente proteina-DNA immagazzina l'energia del legame fosfodiesterico idrolizzato per permettere la reazione inversa al taglio. In quest'ultima fase della reazione è il gruppo OH del filamento interrotto di DNA a operare l'attacco nucleofilo sul legame covalente che unisce la proteina al DNA nell'intermedio di reazione (**fig. 8.6**). Un simile meccanismo d'azione lo abbiamo già incontrato analizzando il meccanismo d'azione delle topoisomerasi del tipo I (*vedi par. 2.4*).

Le cellule e i virus fanno uso della ricombinazione sito-specifica per una varietà di funzioni: in particolare, mediante tale meccanismo molti batteriofagi inseriscono il proprio DNA nel genoma della cellula ospite (fagi λ e P1); in altri casi, è utilizzata per regolare l'espressione genica permettendo, attraverso l'inversione di un tratto di DNA, di esprimere due geni diversi, o serve per mantenere l'integrità strutturale di molecole di DNA circolari durante la loro replicazione.

8.3 Trasposizione

Gli elementi trasponibili (**trasposoni**) sono segmenti di DNA in grado di spostarsi sul DNA da una posizione all'altra, sullo stesso cromosoma o su cromosomi diversi. Queste sequenze mobili di DNA svolgono funzioni importanti: influenzano l'espressione genica, causano mutazioni e riorganizzano la struttura del genoma influenzando profondamente l'evoluzione del materiale genetico.

Gli effetti genetici dei trasposoni furono inizialmente scoperti nel mais da Barbara McClintock nel 1940. Gli organismi vegetali sono particolarmente ricchi di trasposoni e alterazioni dell'espressione genetica nelle piante sono spesso facilmente osservabili come variazioni di colore. I movimenti di un trasposone rendono macchiati i chicchi di una pannocchia di mais saltando dentro e fuori da geni che controllano la loro pigmentazione (**fig. 8.7**). La McClintock notò la capacità dei trasposoni di spostarsi in punti preferenziali sui cromosomi che potevano variare nei diversi ceppi di mais e di trovarsi in posizioni cromosomiche diverse anche nei discendenti di una singola pianta. La comunità scientifica, inizialmente, non apprezzò l'importanza della scoperta della McClintock, ma elementi trasponibili furono successivamente trovati in tutti gli organismi, dai

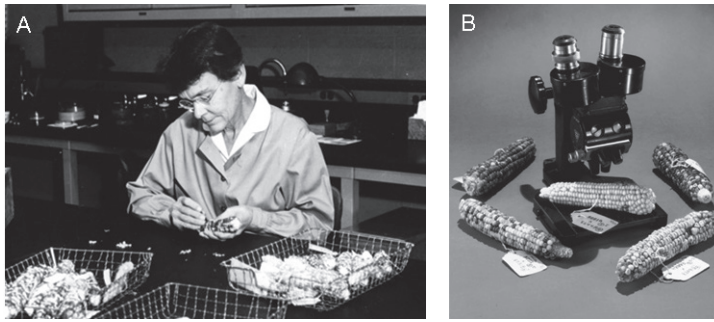


Figura 8.7 Barbara McClintock e gli elementi mobili (trasposoni) del mais. (A) Barbara McClintock (1902-1992); (B) Alcune delle pannocchie di mais utilizzate da Barbara McClintock nei suoi studi sui trasposoni: nei singoli chicchi di una pannocchia di mais è possibile osservare una varietà di pigmentazioni causata da eventi di trasposizione. (Fonte: Smithsonian Institution, National Museum of American History.)

batteri all'uomo. A Barbara McClintock venne assegnato il premio Nobel nel 1983, oltre quarant'anni dopo la loro iniziale scoperta.

Via via che i trasposoni vennero identificati, clonati e sequenziati (*vedi* par. 5.3) fu possibile paragonarli e classificarli in categorie che si differenziano per l'organizzazione generale della loro sequenza e i meccanismi adottati per muoversi sul DNA:

1. trasposoni a DNA;
2. retrotrasposoni, che a loro volta vengono distinti in:
 - retrotrasposoni LTR, a cui appartengono anche i retrovirus, caratterizzati da lunghe sequenze terminali ripetute (chiamate **Long Terminal Repeats** o **LTR**, **finestra 8.3**);
 - retrotrasposoni non-LTR (*vedi* par. 5.3).

Sia i trasposoni a DNA che i retrotrasposoni di tipo retrovirus contengono le sequenze di DNA richieste per la reazione di ricombinazione alla base del meccanismo di trasposizione: nel caso dei trasposoni a DNA sono le sequenze terminali ripetute e invertite, nel caso dei retrotrasposoni di tipo retrovirus le sequenze LTR. Inoltre, entrambi questi trasposoni codificano per l'enzima richiesto per la trasposizione, chiamato **trasposasi** o **integrasi**. I retrotrasposoni di tipo retrovirus codificano anche per la propria trascrittasi inversa (*Reverse Transcriptase* o RT), un tipo speciale di polimerasi in grado di copiare uno stampo di RNA in DNA: la trascrittasi inversa è necessaria in quanto nella trasposizione di questi trasposoni è richiesto un intermedio a RNA (**finestra 8.3**). La struttura dei principali elementi retrotrasponibili è già stata presentata e discussa nel capitolo 5 (*vedi* par. 5.3 e fig. 5.19), così come il fatto che tutti e tre i tipi di trasposoni sopra descritti sono distinti in trasposoni **autonomi** e **non autonomi**. Alla prima classe appartengono i trasposoni che sono in grado di eseguire autonomamente la propria trasposizione: possiedono, infatti, sia le sequenze di DNA in *cis* richieste per la reazione di ricombinazione che le porzioni codificanti per la trasposasi e/o la trascrittasi inversa. Alla seconda classe appartengono quei trasposoni che possono trasporre solo in una cellula dotata di un trasposone autonomo in grado di provvedere la funzione enzimatica richiesta per la reazione di ricombinazione: possiedono, infatti, le sequenze ricombinative in *cis* sul DNA, ma sono privi della loro propria trasposasi.

La tabella 5.6 offre un'idea dell'abbondanza di questi elementi mobili nei vari organismi eucariotici e di come i trasposoni rappresentino una porzione enorme di tutto il genoma. Va però sottolineato che i trasposoni sono presenti non solo negli eucarioti, ma anche nei procarioti.

Senza entrare in dettagli che sono più pertinenti per un testo di Genetica o di Microbiologia, i batteri come *E. coli* possono contenere tre tipi di elementi trasponibili: (1) **sequenze d'inserzione** o **elementi IS**; (2) **trasposoni composti**; (3) **fagi trasponibili**. Gli elementi IS di *E. coli* hanno dimensioni da 600 a 3000 pb e portano alle loro estremità delle sequenze ripetute invertite (IR) di 10-25 pb, mentre la regione centrale codifica per la trasposasi richiesta per la reazione di trasposizione (**fig. 8.8A**). I trasposoni composti di *E. coli* sono più grandi (da 2650 a 9300 pb)

FINESTRA 8.3

I RETROVIRUS

La famiglia *Retroviridae* viene normalmente divisa, basandosi sulla patogenicità, in tre sottofamiglie: (1) **Oncovirus**, che causano trasformazione cellulare *in vivo* e *in vitro*; (2) **Spumavirus**, che causano effetti citopatici *in vitro*, ma non sono stati ancora associati ad alcuna patologia; (3) **Lentivirus**, che inducono lentamente progressive alterazioni del sistema immunitario e danni neurologici. Il virus HIV (Human Immunodeficiency Virus), il più tristemente famoso come causa dell'AIDS nell'uomo, appartiene alla sottofamiglia dei Lentivirus.

Il virione è grossolanamente sferico con un diametro di 100-120 nm e, come risultato dell'infezione di una cellula, i retrovirus (HIV) gemmano dalla membrana plasmatica della cellula infettata (fig. 8F3.1). Ciascun virus contiene all'interno di un capsido proteico due catene a singolo filamento di RNA. La caratteristica peculiare dei retrovirus è la capacità di generare una copia di DNA dal loro genoma a RNA attraverso l'azione di un enzima denominato **trascrittasi inversa**, anche se il nome biochimicamente più appropriato è quello di **DNA polimerasi RNA-dipendente**. La molecola di RNA virale di un tipico retrovirus contiene lunghe ripetizioni terminali (LTR) alle due estremità, ma queste ripetizioni non sono identiche (fig. 8F3.2A). La LTR di sinistra ha una regione R, una regione U5 più una sequenza chiamata PBS o *Primer Binding Site* perché in grado di ibridare con un tRNA che è utilizzato come innescio dalla trascrittasi inversa per sintetizzare DNA fino all'estremità 5' dell'RNA virale (vedi oltre). La LTR di destra contiene una

Figura 8F3.2 Struttura dell'RNA retrovirale e del DNA del provirus e meccanismo di replicazione dell'RNA del retrovirus. (A) Sono da notare le differenze strutturali delle regioni LTR terminali di destra e di sinistra presenti nell'RNA del retrovirus e nel DNA del provirus che viene integrato nel genoma della cellula infettata. (B) Meccanismo di replicazione dell'RNA del retrovirus (vedi il testo per dettagli).

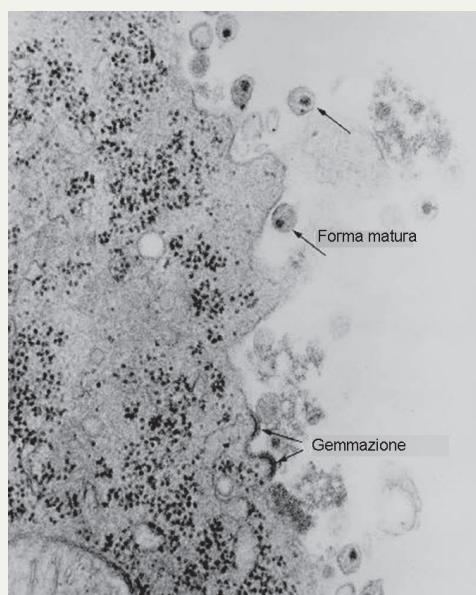
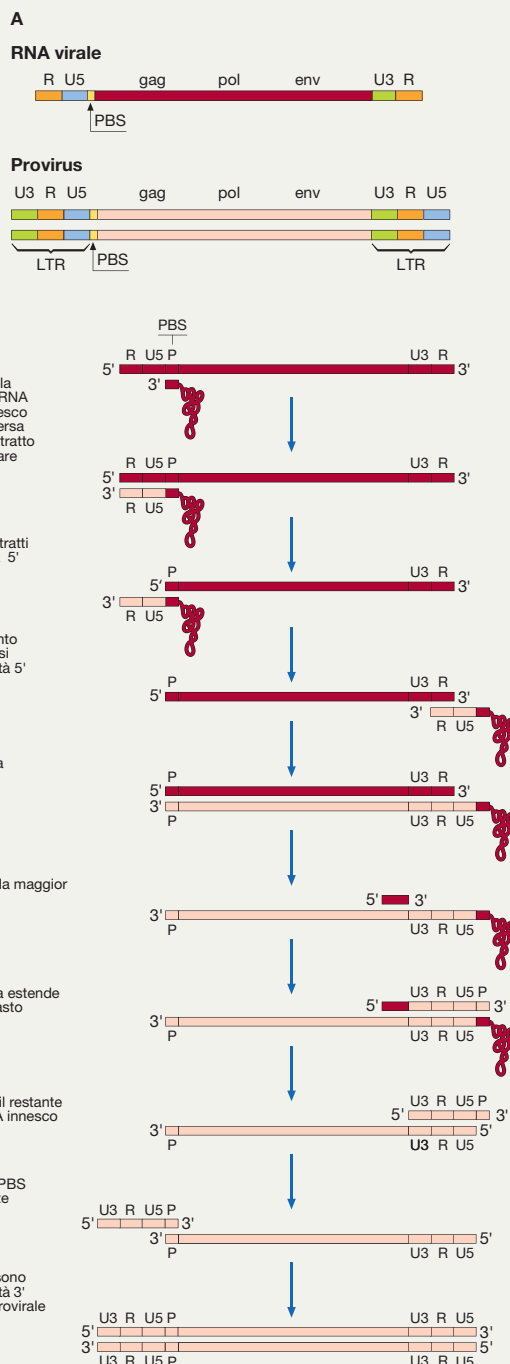


Figura 8F3.1 Una cellula linfocitaria che sta rilasciando particelle virali del virus HIV.

regione U3 e una regione R. Le due estremità dell'RNA virale non sono, quindi, identiche. Al contrario, il DNA provirale che viene prodotto utilizzando l'RNA virale come stampo, contiene LTR complete (U3, R, U5) a ciascuna estremità. Un genoma retrovirale, oltre alle sequenze LTR, contiene tipicamente tre geni chiamati *gag*, *pol* ed *env*. Ciascuna di queste tre regioni codificanti dà origine, in realtà, a più di una proteina per mezzo di reazioni di processamento complesse. Il gene *gag* dà origine a componenti proteici del core nucleoproteico del virione; il gene *pol* codifica per la trascrittasi inversa e per proteine ricombinative; il gene *env* codifica per componenti dell'involucro del virione.

Prima di descrivere come dall'RNA del virus si genera il provirus, è opportuno riassumere il ciclo vitale di un retrovirus (fig. 8F3.3). Un virus infetta una cellula fondendosi con la membrana plasmatica e rilasciando l'RNA virale al suo interno. L'RNA virale (in rosso), come vedremo fra poche righe, dà origine prima a una molecola ibrida RNA:DNA e, successivamente, al DNA a doppio filamento (in verde) corrispondente al provirus che si integra nel genoma della cellula infettata. L'RNA polimerasi II dell'ospite può poi trascrivere il provirus integrato, originando molecole di RNA virale che dopo la loro traduzione possono portare all'assemblaggio di nuove particelle virali mature. Ciò comporta l'impacchettamento dell'RNA in un core nucleoproteico, la sua inclusione in un capsidi di proteine e il distacco di un frammento di membrana dalla cellula ospite durante il processo di gemmazione del virus.

Come si genera il DNA a doppio filamento del provirus a partire dall'RNA virale è descritto in figura 8F3.2B. Come tutte le DNA polimerasi anche la trascrittasi inversa (che è una DNA polimerasi RNA-dipendente) necessita di un innesco. Tale innesco è un tRNA e, precisamente, il tRNA^{Trp} della cellula infettata: quindi, è un componente della cellula stessa a far partire tutto il ciclo di replicazione di un retrovirus! L'allungamento dell'innesco a tRNA da parte della trascrittasi inversa genera un tratto di DNA lungo circa 150 nt che si estende fino all'estremità 5' dell'RNA virale. Cosa succede a questo punto e come è possibile che la replicazione dell'RNA virale generi un provirus le cui estremità sono diverse e più lunghe di quelle dell'RNA virale originario (fig. 8F3.2A)? La risposta a questo paradosso si basa sul fatto che la trascrittasi inversa ha associata un'altra attività, chiamata **RNasi H**, che è in grado di degradare specificamente l'RNA in un ibrido RNA:DNA.

La trascrittasi inversa dopo aver sintetizzato DNA fino all'estremità 5' dell'RNA virale (fase a), utilizza la sua attività RNasi H per degradare le regioni R e U5 dell'estremità sinistra dell'RNA virale (fase b). La rimozione di questo RNA lascia una coda di DNA che, attraverso la sua regione R, può ibridizzare con l'RNA all'altra estremità dello stampo (o con un'altra molecola di RNA stampo) (fase c). L'ibridazione a un'altra regione R è denominata "primo salto" della reazione. Come risultato di questo salto il DNA attaccato all'innesco del tRNA si trova all'estremità destra dello stampo a RNA ed è utilizzato come innesco dalla trascrittasi inversa per copiare la restante parte dell'RNA virale (fase d). Va notato

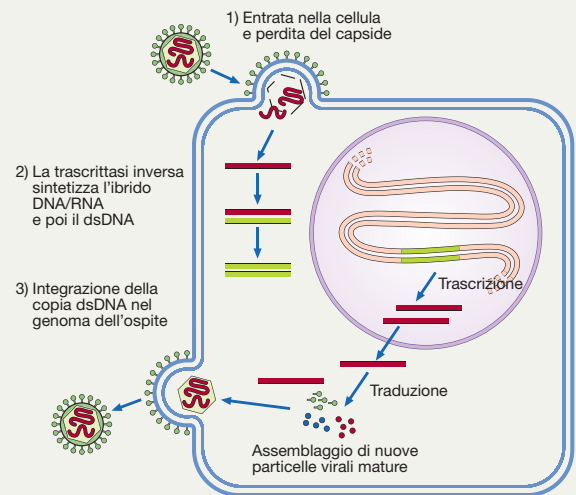


Figura 8F3.3 Ciclo vitale di un retrovirus. Ciascuna particella virale contiene due molecole di RNA di circa 8000-10 000 nucleotidi (in rosso). Tramite l'attività della trascrittasi inversa si genera una molecola ibrida RNA:DNA e, successivamente, una molecola di DNA a doppio filamento (qui indicata in verde per distinguerla dal DNA genomico). Questo DNA copia o provirus viene introdotto nel genoma della cellula infettata mediante l'azione di un enzima, anch'esso codificato dal virus, e chiamato integrasi. Questa integrazione è richiesta per la sintesi, da parte dell'RNA polimerasi della cellula, di nuove molecole di RNA virali.

che il primo salto ha permesso il completamento dell'LTR di destra. Successivamente, l'attività RNasi H rimuove gran parte dell'RNA virale lasciando una piccola porzione di RNA nelle vicinanze dell'LTR di destra (fase e) che serve come innesco per la sintesi del secondo filamento di DNA (fase f). Dopo che la trascrittasi inversa ha allungato lo stampo fino all'estremità di destra, l'RNasi H rimuove l'RNA rimasto, cioè l'innesco del secondo filamento e il tRNA (fase g) associati al DNA. Ciò porta al secondo salto della reazione (fase h), in cui la regione P sulla destra si appaia alla regione P di sinistra. Come per il primo salto, anche il secondo può essere immaginato come un salto su un'altra molecola, o sull'altra estremità della stessa molecola. In quest'ultimo caso, il DNA può avvolgersi in modo tale da permettere alle due regioni di appaiarsi. Dopo il secondo salto, una DNA polimerasi (o forse la trascrittasi inversa stessa) completa entrambi i filamenti (fase i), utilizzando come stampo le sporgenze di DNA a singolo filamento su ciascuna estremità.

Una volta che il DNA del provirus è stato sintetizzato, può essere inserito nel genoma dell'ospite mediante una **integrasi**. Questo enzima fa parte inizialmente della **poliproteina** prodotta dal gene *pol*, che codifica anche la trascrittasi inversa e l'RNasi H. L'integrasi è tagliata dalla poliproteina mediante l'azione di una **proteasi**, che inizialmente fa anch'essa parte della stessa poliproteina. È interessante notare che alcuni dei farmaci più promettenti nella cura dell'AIDS sono inibitori di proteasi che hanno come bersaglio l'enzima specifico codificato dal virus dell'HIV.

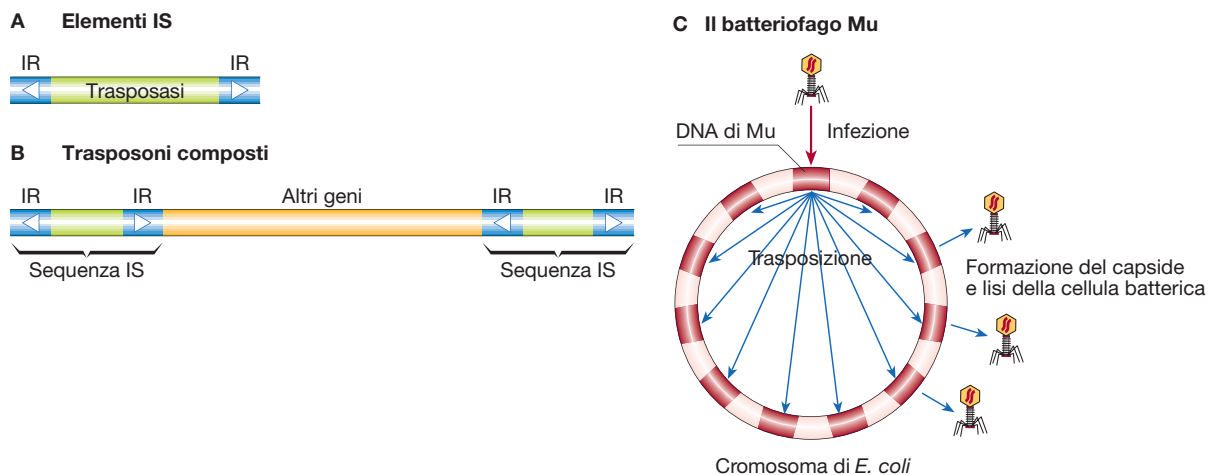


Figura 8.8 Elementi genetici mobili nel batterio *Escherichia coli*. La struttura degli elementi trasponibili di *E. coli* è descritta nel testo: (A) sequenze di inserzione IS; (B) trasposoni composti Tn; (C) fago trasponibile Mu.

perché, oltre ai geni per la trasposizione, contengono altre funzioni, tra cui geni che portano alla resistenza dei batteri al trattamento con alcuni antibiotici (**fig. 8.8B**). Il batteriofago Mu è un virus batterico temperato che si replica con il meccanismo della trasposizione. A differenza, però, degli altri elementi genetici mobili, il DNA del fago Mu può esistere come unità extracromosomica e venire impacchettato in un capsido, potendo, così, infettare cellule batteriche. Quando Mu infetta un batterio, il suo DNA lineare circolarizza e si integra nel DNA circolare di *E. coli*. Successivamente, in diverse condizioni il DNA di Mu può trasporre in numerose posizioni sul cromosoma di *E. coli*. Le diverse copie di Mu così generate possono poi venire impacchettate nel capsido e generare la lisi della cellula (**fig. 8.8C**).

Sia le sequenze IS che i trasposoni composti hanno brevi sequenze ripetute invertite all'estremità del loro DNA. Queste sequenze sono coinvolte nel processo di trasposizione. Quando un elemento trasponibile si inserisce in un altro punto del genoma di *E. coli*, si verifica la duplicazione di una breve sequenza del DNA bersaglio a livello del sito d'integrazione (**fig. 8.9**). Questa sequenza bersaglio non era presente nel trasposone, ma è l'elemento trasponibile a determinare la duplicazione di questo DNA attraverso il processo d'inserzione. La duplicazione della sequenza bersaglio si verifica in seguito all'attività dell'enzima trasposasi che induce tagli sfalsati sul DNA. Il trasposone si attacca alle estremità a singolo filamento generate e la loro riparazione determina la duplicazione. Alcuni elementi trasponibili hanno delle sequenze bersaglio preferenziali, mentre altri, incluso il batteriofago Mu, possono inserirsi in modo quasi del tutto casuale nel genoma.

I meccanismi coinvolti nella trasposizione sono diversi per i vari trasposoni di *E. coli*. Due elementi sono, comunque, comuni in tutti i casi: il coinvolgimento delle sequenze ripetute invertite e la necessità di una trasposasi attiva. Esistono poi due alternative: (1) un procedimento conservativo, denominato "taglia e cuci", in cui l'elemento trasponibile viene tagliato via dal sito donatore e inserito nel sito accettore (tale meccanismo si verifica per il trasposone Tn5 di *E. coli*); (2) un meccanismo di trasposizione che passa attraverso una fase di replicazione del DNA. In quest'ultimo caso, la trasposasi determina la formazione di un legame temporaneo fra DNA donatore e DNA accettore formando un intermedio chiamato **co-integrato** contenente due copie del trasposone. La risoluzione di tale intermedio fa sì che una copia del trasposone rimanga nel sito di origine, mentre la seconda copia viene integrata nella nuova sede (**fig. 8.10**). Il trasposone Tn3 di *E. coli* viene trasposto utilizzando il meccanismo replicativo, mentre il

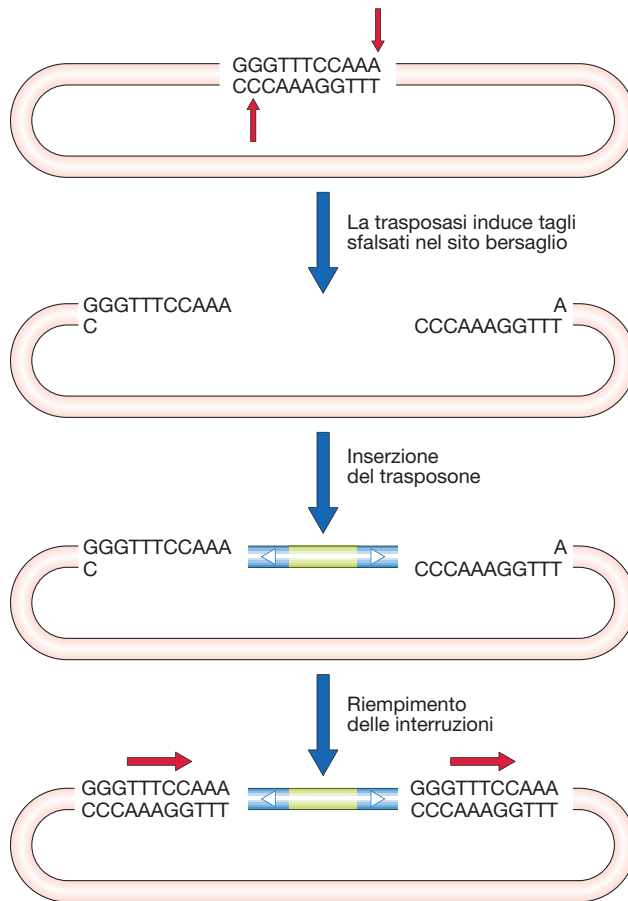


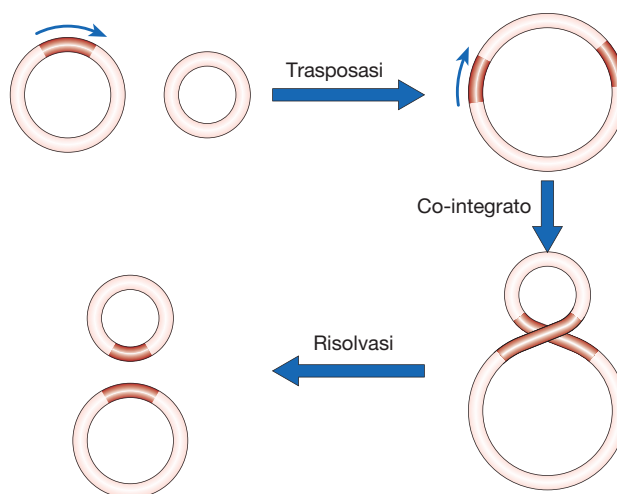
Figura 8.9 La trasposizione di un elemento IS o di un trasposone composto genera la duplicazione di una sequenza di DNA. Come descritto nel testo, in seguito al meccanismo d'azione della trasposasi, si genera la duplicazione di una breve sequenza del DNA bersaglio a livello del sito d'integrazione del trasposone.

DNA del fago Mu può trasporre secondo entrambi i meccanismi descritti. Al di là dell'evidente importanza che hanno avuto e hanno i trasposoni nell'evoluzione dei genomi, qui di seguito descriveremo pochi altri esempi di trasposizione negli eucarioti per sottolinearne, da una parte, la valenza quantitativa e, dall'altra, la rilevanza nel modulare specifiche funzioni.

Della rilevanza e quantità dei LINE (*Long Interspersed Elements*) e dei SINE (*Short Interspersed Element*) nel genoma umano abbiamo già parlato nel paragrafo 5.3. In alcuni organismi, i trasposoni hanno un'enorme importanza nel mantenimento della struttura e nella trasmissione dei cromosomi. Nel moscerino della frutta *Drosophila melanogaster*, i trasposoni sono importanti per il mantenimento dei telomeri. I telomeri costituiscono le estremità dei cromosomi e sono caratterizzati dalla ripetizione, anche per centinaia di volte, di una corta sequenza di DNA conservata nel corso dell'evoluzione (5'-TTAGGG-3' nei vertebrati). Queste ripetizioni di sequenza ai telomeri sono assicurate dal meccanismo di replicazione del DNA telomerico (vedi par. 6.6), che richiede l'intervento di un particolare enzima, chiamato "telomerasi", che, come abbiamo già descritto, è una trascrittasi inversa un po' particolare. *Drosophila*, a un certo punto della sua storia evolutiva, ha perso il gene per la telomerasi e i telomeri delle sue quattro copie di cromosomi sono costituiti da due retrotrasposoni poli-A, chiamati *HeT-A* e *TART*. Questi trasposoni sono correlati ad altri elementi trasponibili presenti altrove nel genoma di *Drosophila*, ma per ragioni non ancora completamente chiarite traspongono solo in corrispondenza dei telomeri, contribuendo al mantenimento della loro corretta dimensione.

Recentemente, è stato trovato che i centromeri di mais e di molti cereali sono costituiti da ripetizioni di sequenze di DNA, tra cui il retrotrasposone CRM.

Figura 8.10 Trasposizione attraverso un processo replicativo che prevede la formazione di un co-integrato. Vedi il testo per dettagli.



I centromeri identificano regioni di DNA nei cromosomi su cui si assemblano una serie di proteine fino a costituire i cinetocori, strutture nucleo-proteiche specializzate a cui si attaccano le fibre del fuso richieste per la corretta segregazione dei cromosomi in mitosi. La proteina CENH3 è simile all'istone H3 e lo sostituisce nella cromatina centromerica. Mediante esperimenti di ChIP (*vedi* par. 20.12) è stato trovato che CENH3 interagisce con la sequenza del retrotrasposone CRM presente solo in corrispondenza delle regioni centromeriche. Un altro punto importante che collega i trasposoni con i centromeri cromosomici è dato dal fatto che la proteina CENP-B, un fattore essenziale del cinetocore, mostra una sorprendente similarità di sequenza con una trasposasi suggerendo una relazione di omologia, ovvero una comune origine evolutiva (*vedi* par. 4.6). La presenza preferenziale di trasposoni nelle regioni centromeriche o il ruolo di proteine trasposasi-simili nel cinetocore sono, però, ancora da definire.

8.4 Riarrangiamenti di sequenze di DNA e controllo dell'espressione genica

Descriviamo in questo paragrafo alcune situazioni in cui cambiamenti qualitativi (riarrangiamenti) nell'organizzazione di sequenze genomiche modificano l'espressione genica sia in procarioti che in eucarioti.

Cambiamento della specificità d'ospite nel fago Mu

Come descritto poco sopra, il fago Mu è un virus che infetta *Escherichia coli*, integra il suo genoma casualmente nel cromosoma dell'ospite e si replica e si sposta come un trasposone. Come mostrato in **figura 8.11A** il genoma del fago Mu comprende una regione di circa 3 Kpb chiamata **segmento G**, o **segmento invertibile**, che, infatti, si trova in orientamento opposto in diverse molecole di DNA di Mu. Questa regione contiene sequenze che codificano per proteine implicate nell'assorbimento del virus sulla cellula ospite. In particolare il gene S codifica per le fibre della coda del fago che sono responsabili del riconoscimento della cellula ospite e, quindi, della specificità d'ospite del virus. L'inversione del frammento G provoca un cambiamento della proteina S sintetizzata e quindi un cambiamento della specificità di ospite, che può essere *E. coli* K12 oppure *E. coli* C.

Andando un po' più nel dettaglio (**fig. 8.11B-C**), osserviamo che il prodotto genico di S comprende una regione costante (Sc) e una regione variabile di cui

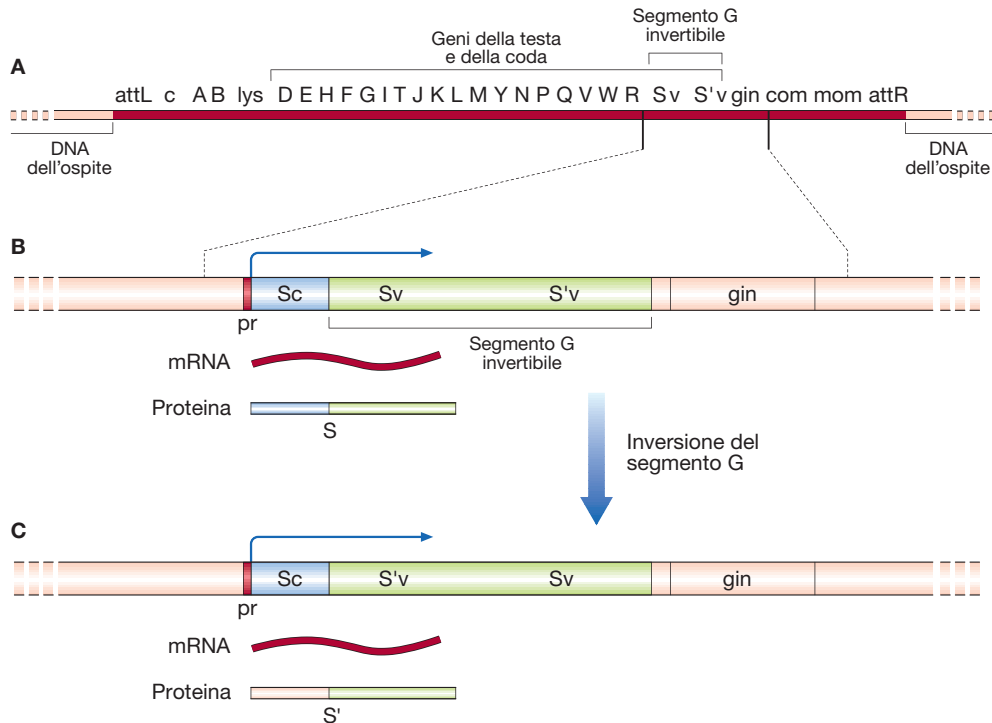


Figura 8.11 L'inversione del segmento G del cromosoma del fago Mu determina il cambiamento della specificità d'ospite. (A) Mappa genetica del fago Mu. (B) Dettaglio della mappa genetica: il frammento G invertibile è orientato in modo che a valle di Sc si trova la regione Sv, risultando nella sintesi della proteina S. (C) L'inversione del frammento invertibile G, promossa dal prodotto del gene *gin*, porta la regione S'v a valle di Sc e risulta quindi nella sintesi della proteina S'.

esistono due forme (Sv e S'v). La regione codificante per la porzione costante Sc è preceduta da un promotore e si trova subito a monte del segmento G. La porzione variabile è codificata da una delle due sequenze Sv o S'v che si trovano agli estremi del segmento G invertibile, cosicché l'inversione risulta in un cambiamento della sequenza (Sv o S'v) che si viene a trovare a valle di Sc. Ovviamente le due regioni codificanti Sv e S'v si devono trovare sui due diversi filamenti del DNA in modo che, con l'inversione, la ORF Sv (o S'v) si venga a trovare sullo stesso filamento di quella di Sc. Inoltre, è importante che l'evento di ricombinazione sito-specifica che porta all'inversione sia estremamente preciso, al singolo nucleotide, in modo che venga mantenuta la ORF per l'intera proteina. L'enzima responsabile dell'inversione del segmento G è conosciuto come Gin (**G invertasi**) ed è codificato dal gene fagico *gin*, che, nella mappa genetica di **figura 8.11**, si trova alla destra del segmento G. L'enzima e il sistema di ricombinazione sito-specifica implicati nell'inversione del segmento G presentano notevoli affinità con la risolvasi e il sistema di trasposizione del trasposone Tn3.

Variazione di fase in *Salmonella*

Un sistema analogo a quello sopra descritto è responsabile, in *Salmonella*, del **fenomeno della variazione di fase**. Questo processo, presente anche in altri tipi di batteri, consiste in un cambiamento degli antigeni di superficie del batterio che così si sottrae alla risposta immunitaria specifica innescata dall'ospite. In particolare *Salmonella* può esistere in una di due possibili fasi.

In fase 1 le cellule sono molto mobili grazie all'attività di flagelli costituiti dalla proteina H1 (chiamata anche **flagellina 1**). Nella fase 2 le cellule poco mobili non esprimono H1 ma H2 (o **flagellina 2**). Come già detto, il cambiamento

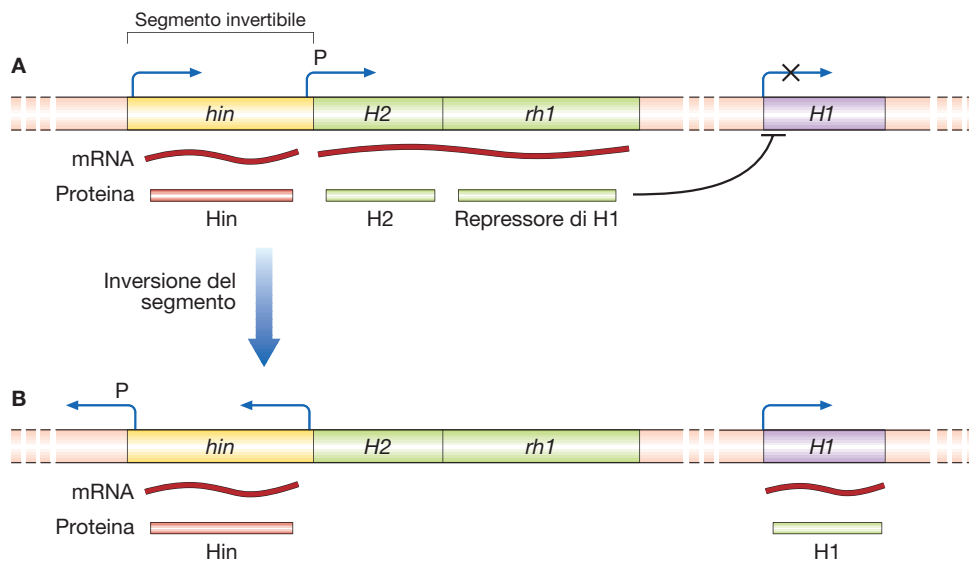


Figura 8.12 L'inversione di un segmento cromosomico è alla base della variazione di fase in *Salmonella*. (A) Nella *Salmonella* in fase 2 il promotore *P* promuove la trascrizione dei geni *H2*, per la flagellina 2, e *rh1*, per un repressore trascrizionale che reprime la trascrizione del gene *H1*. (B) L'inversione del frammento invertibile,

promossa dal prodotto del gene *hin*, sposta il promotore *P* e lo orienta verso l'adiacente regione cromosomica dell'ospite. Conseguentemente i geni *H2* e *rh1* non sono più trascritti, mentre viene trascritto il gene *H1* non più represso. Il gene *hin* esprime la proteina Hin necessaria per le inversioni, in ambedue gli orientamenti.

di fase permette alla cellula di sottrarsi alle difese immunitarie dell'ospite. Come mostrato in **figura 8.12**, *Salmonella* ha due geni che codificano per le due diverse forme di flagellina, *H1* e *H2*. Il gene *H2* è seguito a valle dal gene *rh1* che codifica per un repressore della trascrizione di *H1*; i due geni sono trascritti in un unico mRNA. A monte di *H2* si trova un segmento di DNA invertibile, lungo circa 1000 pb, che contiene il gene *hin*. Questo codifica per la proteina Hin, che, analogamente a Gin del fago Mu, media l'inversione del frammento stesso, il che avviene con una frequenza di 1 volta ogni circa 1000 divisioni batteriche. Il frammento invertibile presenta due promotori: uno di questi orienta la trascrizione verso l'interno del segmento ed è infatti il promotore del gene *hin*; l'altro promotore (indicato con *P* in figura) si trova all'altra estremità del segmento invertibile ed è orientato in modo da avviare la trascrizione verso la sequenza genomica adiacente.

Infine, c'è il gene *H1* per una diversa forma di flagellina che si trova localizzato in un'altra regione del cromosoma ed è dotato di un suo promotore controllato negativamente dal repressore sopra menzionato. Abbiamo descritto l'organizzazione di questa regione genomica, vediamo ora come funziona.

Se l'orientamento del segmento invertibile è quello mostrato in **figura 8.12A**, il promotore *P* promuove la trascrizione dei geni *H2* e *rh1*, portando alla sintesi della flagellina *H2* e del repressore che va a bloccare la trascrizione del gene *H1*. In queste condizioni il batterio è in fase 2. Come mostrato in **figura 8.12B**, l'inversione del frammento, mediata da Hin, fa sì che il promotore *P* non trasciva più i geni *H2* e *rh1* (potrà eventualmente promuovere la trascrizione di sequenze genomiche adiacenti non rilevanti). Comunque, l'assenza del repressore permette l'avvio della trascrizione di *H1* a partire dal suo promotore, e il batterio risulterà in fase 1.

Cambiamento del tipo sessuale in lievito (locus *MAT*)

Come verrà descritto anche nel paragrafo 22.2, il lievito *Saccharomyces cerevisiae* può crescere sia in uno stato aploide che in uno stato diploide. I ceppi aploidi



Figura 8.13 Ira Herskowitz (1946-2003). Il genetista Ira Herskowitz, dopo aver ottenuto importanti risultati nello studio della regolazione del fago lambda, è passato al sistema di lievito in cui ha tra l'altro contribuito in maniera fondamentale allo studio del locus MAT.

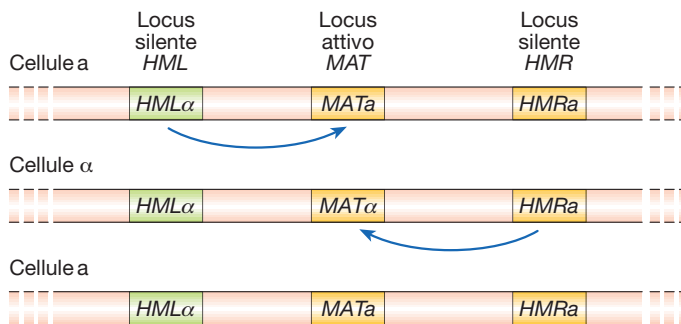


Figura 8.14 Modello "a cassette" per il cambiamento di tipo sessuale nel lievito *Saccharomyces cerevisiae*. La sequenza genica a (o α) presente nel locus MAT è trascrizionalmente attiva, mentre le sequenze a e α presenti rispettivamente in HML e HMR non sono espresse. Quando una sequenza α di HML α va a sostituire la sequenza a nel locus MAT, la cellula cambia tipo sessuale da a ad α . Analogamente, quando una sequenza a di HMR a va a sostituire la sequenza α nel locus MAT, la cellula cambia tipo sessuale da α ad a .

possono essere di due tipi sessuali a e α (vedi fig. 22.2). Due cellule aploidi di diverso tipo sessuale, una a e una α , possono accoppiarsi mediante un processo di coniugazione producendo una forma diploide a/α . Questa può moltiplicarsi come diploide e può eventualmente andare incontro a un evento meiotico producendo quattro cellule aploidi, due di tipo a e due di tipo α . Un'osservazione che incuriosì molto i ricercatori è che tra la progenie di cellule a compaiono occasionalmente anche cellule α e, viceversa, tra la progenie delle cellule α compaiono occasionalmente anche cellule a . Questo cambiamento di tipo sessuale è dovuto a un riarrangiamento delle sequenze genomiche che è stato ben studiato e caratterizzato da numerosi ricercatori tra cui menzioniamo Ira Herskowitz (fig. 8.13). Noi ci limiteremo a descrivere il sistema nelle sue linee generali.

La figura 8.14 mostra la regione cromosomica implicata in questo fenomeno. Al centro di questa regione c'è un locus chiamato **MAT** (*Mating Type*, tipo sessuale) trascrizionalmente attivo, che codifica per fattori trascrizionali che regolano l'espressione di geni per funzioni coinvolte nell'accoppiamento, quali i feromoni e i loro recettori. A seconda che nel locus MAT sia presente l'allele *MATa* o l'allele *MATα* la cellula sarà di tipo sessuale a o α . A una certa distanza sullo stesso cromosoma, vi sono altri due loci, *HML* e *HMR*, a sinistra e a destra rispetto a MAT nella figura. Questi contengono, rispettivamente, una copia del gene α (*HMLα*) e una copia del gene a (*HMRa*), copie geniche che però, fintanto che sono localizzate nei loci *HML* e *HMR*, non sono trascrizionalmente attive. Occasionalmente, mediante un meccanismo che vedremo più avanti, una di queste due sequenze va a sostituire la sequenza presente nel locus MAT. Nell'esempio mostrato in figura 8.14 il locus MAT contiene inizialmente l'allele a ; la sostituzione di questo con la sequenza α provoca un cambiamento del tipo

sessuale della cellula da a ad α . Una successiva occasionale sostituzione della sequenza a al posto della α cambierà nuovamente il tipo sessuale, da α ad a . Pertanto, da qualsiasi condizione iniziamo, a o α , dopo qualche generazione saranno presenti nella popolazione ambedue i tipi sessuali. Può anche succedere che una sequenza a vada a sostituire una a già presente nel sito attivo, o che una α vada a sostituire una α . In questi casi non ci sarà cambiamento del tipo sessuale, anzi, la sostituzione non sarà nemmeno sperimentalmente apprezzabile.

Il modello mostrato in **figura 8.14** è anche chiamato **modello a cassette** per analogia con un mangiacassette (mangianastri) in cui si può sostituire la cassetta (nastro) che si vuole ascoltare. *HMR α* e *HMR a* corrispondono a due cassette inattive che vengono attivate quando inserite nel mangianastri *MAT*.

Qual è il meccanismo di riarrangiamento genomico che risulta nella sostituzione della sequenza che si trova nel locus *MAT*? È stato dimostrato che qui il meccanismo è simile a quello implicato nei fenomeni di “conversione genica” (vedi finestra 8.1). Il modello oggi accettato è illustrato in **figura 8.15** che prende in considerazione, come esempio, la sostituzione di α nel locus *MAT* (*MAT α*) con la sequenza a che si trova in *HMR a* . Il processo inizia con un taglio della doppia elica, vicino a un'estremità del locus *MAT*, effettuato da una nucleasi nota come HO. Questa è un'endonucleasi altamente specifica che riconosce una sequenza bersaglio presente, nel genoma di lievito, solo in questo sito del locus

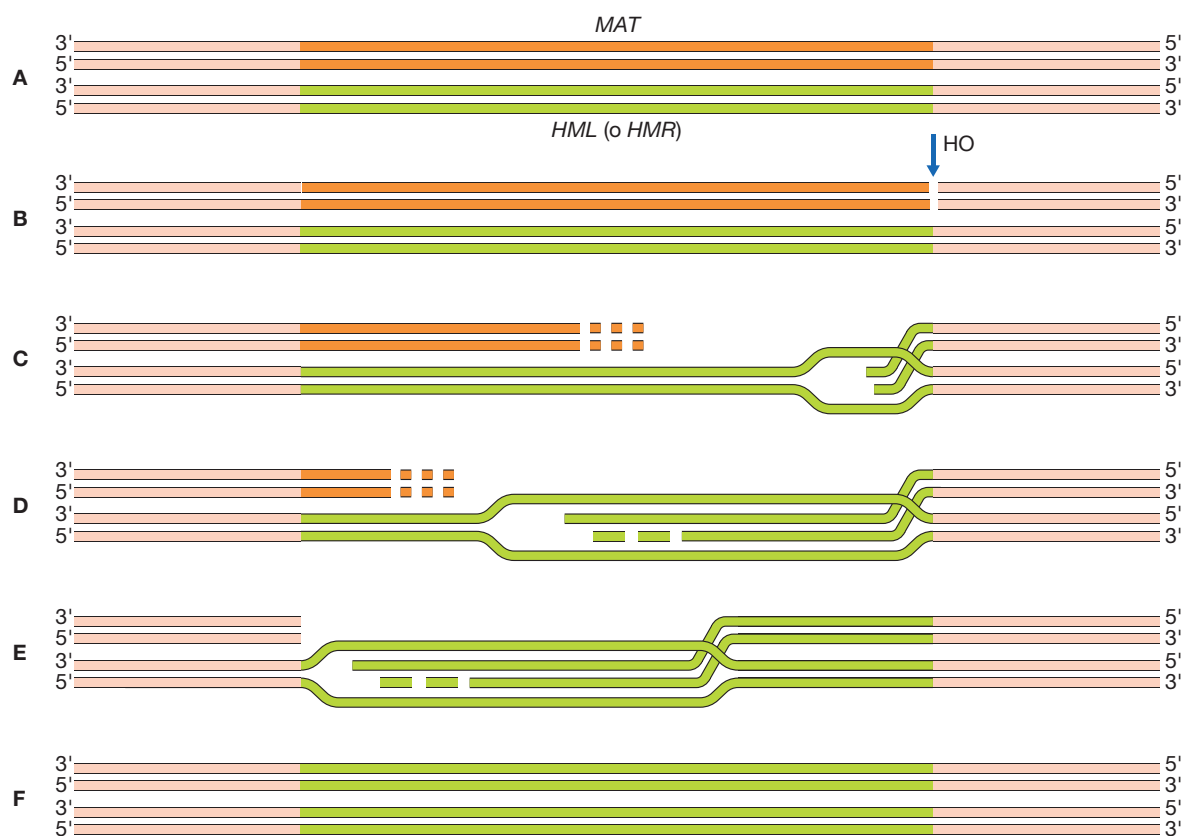


Figura 8.15 La sostituzione della sequenza di *MAT* con la sequenza di *HML* (o *HMR*) avviene tramite un meccanismo di conversione genica. (A) Appaiamento delle sequenze del locus ricevente *MAT* (arancione) e del locus donatore *HML* o *HMR* (verde). (B) L'endonucleasi HO provoca un taglio a doppio filamento in prossimità dell'estremità di *MAT*. (C) *MAT* viene degradato a

partire dal sito di taglio, mentre le altre estremità generate dal taglio "invadono" la doppia elica all'estremità di *HML* (o *HMR*) appaiandosi ai filamenti che si sono separati. (D) ed (E) La forcella di replicazione che si è assemblata avanza sintetizzando due nuovi filamenti, uno continuo l'altro discontinuo. (F) La ligasi ristabilisce la continuità della doppia elica del locus *MAT*.

MAT, dove provoca un taglio sfalsato del doppio filamento. Il DNA del locus *MAT* viene degradato e le estremità dei due filamenti tagliati vanno ad appaiarsi con i due filamenti separati della doppia elica del locus *HMRA*, che gli si è affiancato. Si crea così una tipica forcella di replicazione su cui si assembla l'apparato di replicazione che procede, verso sinistra nella figura, replicando ambedue i filamenti, un filamento continuo e uno discontinuo. Mano a mano che la forcella di replicazione avanza, i filamenti di nuova sintesi si dissociano dai parentali e vanno a riformare una doppia elica fino a quando, raggiunta l'estremità sinistra del locus, la ligasi provvederà a ricostituire l'integrità della doppia elica del locus *MAT*, che a questo punto non è più *MAT α* ma *MAT a* .

Variazione antigenica in *Trypanosoma* (geni VSG)

I tripanosomi hanno un ciclo vitale complesso che si svolge alternativamente in due ospiti, spesso un insetto e un mammifero, per esempio la mosca *tse-tse* e l'uomo, nel quale il tripanosoma provoca la malattia del sonno. Durante la fase di proliferazione nel mammifero il tripanosoma è rivestito, al di sopra della membrana plasmatica, da un monostrato costituito da qualche milione di molecole di una singola glicoproteina chiamata **VSG** (*Variable Surface Glycoprotein*, glicoproteina variabile di superficie) che rappresenta l'unico antigene esposto all'esterno dal tripanosoma nell'ospite umano. Dopo circa un paio di settimane, quando il sistema immunitario dell'individuo ospite inizia a produrre gli anticorpi specifici contro la VSG del tripanosoma, questo inizia a sintetizzare una nuova VSG, immunologicamente diversa dalla precedente, sottraendosi così alle difese dell'ospite. Si noterà un'analogia tra questa transizione, chiamata **variazione antigenica**, e la variazione di fase che abbiamo visto avvenire in *Salmonella*; con la differenza che in *Salmonella* l'inversione del frammento genomico permette l'alternanza di due soli antigeni di superficie, mentre in *Trypanosoma* le possibilità sono di più. Infatti, *Trypanosoma* può sintetizzare decine di diverse VSG, sempre una alla volta, con un ordine variabile non predeterminato. Questo fenomeno spiega le continue ricadute e febbri ricorrenti caratteristiche di questo tipo di malattie e le difficoltà incontrate dai pazienti a superare l'infezione.

Qual è il meccanismo molecolare alla base di questa variazione antigenica? Anche in questo caso sono coinvolti riarrangiamenti delle sequenze genomiche, ma più complessi di quelli visti fino ad ora. Come abbiamo detto, oltre a una variazione dell'antigene di superficie, il sistema garantisce anche che l'ordine con cui vengono espresse le varie VSG non sia determinato, ma puramente casuale. Questo è il risultato del contemporaneo intervento di due processi che sono illustrati in **figura 8.16** con riferimento alla specie *Trypanosoma brucei*. Questa specie, infatti, cresce bene anche in animali da laboratorio e rappresenta quindi il sistema meglio studiato. Va precisato che lo schema che presentiamo è alquanto semplificato e non entra nel dettaglio di meccanismi molecolari che sono stati in parte chiariti. Le numerose copie di sequenze geniche VSG possono essere distinte in due categorie. Alcune copie si trovano localizzate in regioni cromosomiche adiacenti ai telomeri di alcuni dei numerosi cromosomi del tripanosoma. Solo alcuni di questi siti telomerici possono essere attivati trascrizionalmente e produrre la VSG, ma non più di uno alla volta. La figura mostra tre di questi siti, *A*, *B* e *C*, uno dei quali, *A* in questo caso, è trascrizionalmente attivo ed è chiamato ELC (*Expression-Linked Copy*, copia associata all'espressione). Gli altri due siti, *B* e *C*, potenzialmente attivi, sono al momento silenti. Se uno di questi due siti, per esempio *C*, viene attivato, il sito *A* precedentemente attivo viene silenziato. La successione con cui sono attivati e disattivati questi siti telomerici non è determinata ma casuale. Questo processo è in grado di produrre una certa variazione antigenica, ma limitata a poche varianti di VSG, tre nel nostro esempio. Esistono poi nel genoma numerose altre copie di sequenze VSG, anche

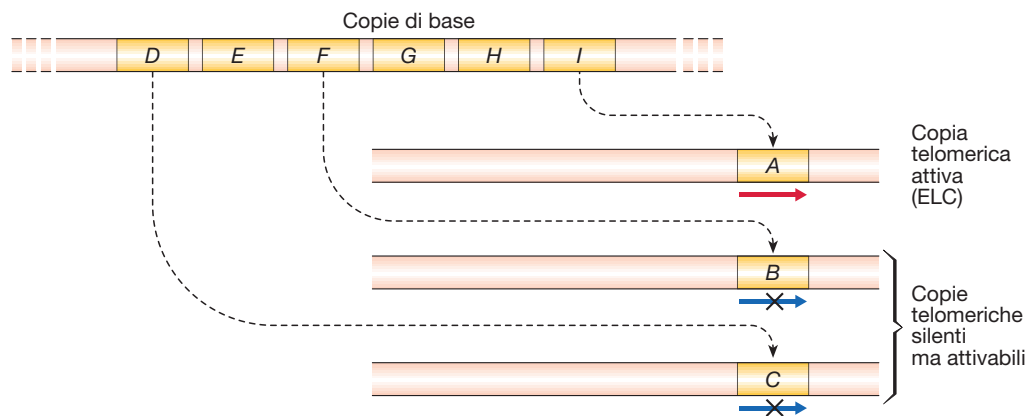


Figura 8.16 I geni VSG responsabili della variazione antigenica dei tripanosomi. Alcune copie telomeriche dei geni VSG possono essere attivate trascrizionalmente, ma solo una alla volta. In questo esempio è espressa la copia A. Se viene attivata una delle altre due copie telomeriche, B o C, la copia A viene silenziata.

Numerose copie di base (D-I in questo esempio) sono raggruppate in loci interni di alcuni cromosomi dove non vengono mai espresse. Occasionalmente, queste vanno a sostituire le copie telomeriche dove, eventualmente, possono essere attivate trascrizionalmente.

centinaia, organizzate per lo più in tandem a formare gruppi localizzati in regioni interne dei cromosomi.

In **figura 8.16** è mostrato per semplicità un solo gruppo di sei copie, D-I. Questi geni, chiamati **geni di base**, non sono attivi. Occasionalmente queste copie possono andare a sostituire una delle copie telomeriche e diventare quindi attivabili e, se quel sito telomerico viene attivato, diventare la copia effettivamente attiva ELC che produce la nuova VSG. Per quanto riguarda il meccanismo molecolare che permette questa sostituzione di sequenze nei loci telomerici attivabili, sembra essere un meccanismo analogo alla conversione genica, che abbiamo descritto per il locus *MAT* di lievito. Abbiamo detto che la variazione antigenica del tripanosoma assomiglia, dal punto di vista della funzione, alla variazione di fase di *Salmonella*. Possiamo ora dire che invece, dal punto di vista del meccanismo molecolare, la variazione antigenica assomiglia al locus *MAT* di lievito; con la differenza che, mentre per *MAT* di lievito ci sono solo due cassette da alternare nel sito attivo, nel sistema VSG di tripanosoma ce ne sono decine o anche centinaia. Come il meccanismo del cambiamento di tipo sessuale del lievito è stato chiamato **modello a cassette**, potremmo paragonare il meccanismo alla base della variazione antigenica di tripanosoma a un DJ (*disk-jockey*) con la sua console. Infatti, il DJ ha a disposizione due o tre giradischi che utilizza uno alla volta alternandoli e una collezione di dischi che va continuamente a sostituire sui tre giradischi.

Ricombinazione V(D)J dei geni per le immunoglobuline nei vertebrati

Si arriva così al **sistema di ricombinazione V(D)J**, responsabile dell'assemblaggio di frammenti di geni durante lo sviluppo del sistema immunitario dei vertebrati che, di tutti i meccanismi di differenziamento cellulare basati su fenomeni di riarrangiamenti genomici, è di gran lunga il più complesso e interessante. Non è nostro scopo descrivere e discutere qui la risposta immunitaria e i suoi meccanismi, anche perché l'argomento è in genere oggetto di un intero corso. Presenteremo solo una breve e schematica presentazione del problema per completare il discorso sui meccanismi di regolazione che coinvolgono riarrangiamenti genomici e vogliamo sottolineare che, nel caso della ricombinazione V(D)J,

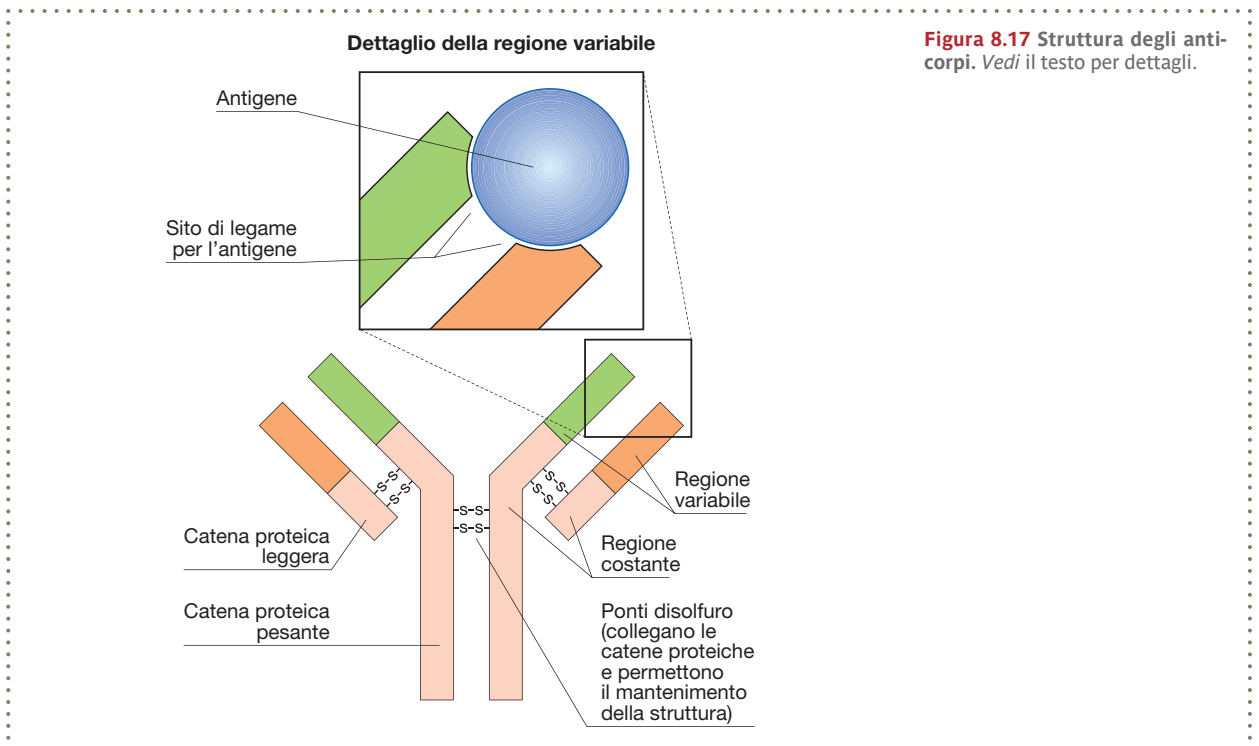


Figura 8.17 Struttura degli anticorpi. Vedi il testo per dettagli.

il meccanismo di riarrangiamento è sostanzialmente quello della trasposizione, come per la variazione di fase di *Salmonella* e la variazione di specificità d'ospite del fago Mu, a differenza del cambiamento di tipo sessuale di lievito e della variazione antigenica di *Trypanosoma* che, come abbiamo visto, utilizzano meccanismi analoghi alla conversione genica.

Brevemente, i vertebrati si difendono dall'attacco di virus, batteri e altri agenti patogeni tramite il funzionamento del loro sistema immunitario, che è costituito principalmente dai linfociti B e T. Le cellule B producono gli anticorpi solubili presenti nel circolo sanguigno, mentre le cellule T sintetizzano dei recettori presenti sulla superficie cellulare (detti "recettori delle cellule T"). Gli anticorpi solubili e i recettori delle cellule T devono essere in grado di riconoscere un'enorme varietà di molecole (antigeni) presenti su organismi patogeni per identificarli e innescare la risposta immunitaria che porta alla loro eliminazione. Il meccanismo utilizzato dalle cellule per produrre tale enorme varietà di anticorpi e recettori delle cellule T consiste in una serie specializzata di riarrangiamenti sul DNA, chiamata **ricombinazione V(D)J**. La natura di tali riarrangiamenti fu scoperta quando furono clonati e sequenziati i primi geni codificanti per anticorpi.

Ogni molecola di anticorpo è costituita da 4 polipeptidi tenuti insieme da ponti disolfuro: due catene pesanti (H) più grandi e uguali tra di loro, e due catene leggere (L) più piccole e uguali tra di loro (**fig. 8.17**). Ogni catena è poi costituita da una regione costante (C), che è essenzialmente condivisa in tutti gli anticorpi, e da una regione variabile (V), che è specifica per ciascun anticorpo. Nel gene codificante per la catena leggera (V_k) nel topo ci sono circa 300 regioni V ($V_1, V_2, V_3 \dots V_n$) che codificano per versioni diverse della regione variabile della catena leggera (**fig. 8.18**). Tra la regione V e l'unica regione costante C, ci sono 4 segmenti genici (chiamati J, per *Joining*). Ne deriva che tramite ricombinazione tra segmenti V e J si possono generare 1200 versioni diverse di catene leggere per un anticorpo. La costituzione della catena pesante di un anticorpo è molto simile: in questo caso, però, un ulteriore segmento del gene (denominato D, per *Diversity*) è coinvolto nella generazione dei diversi tipi di anticorpo. La situazione può essere molto complessa: per esempio,

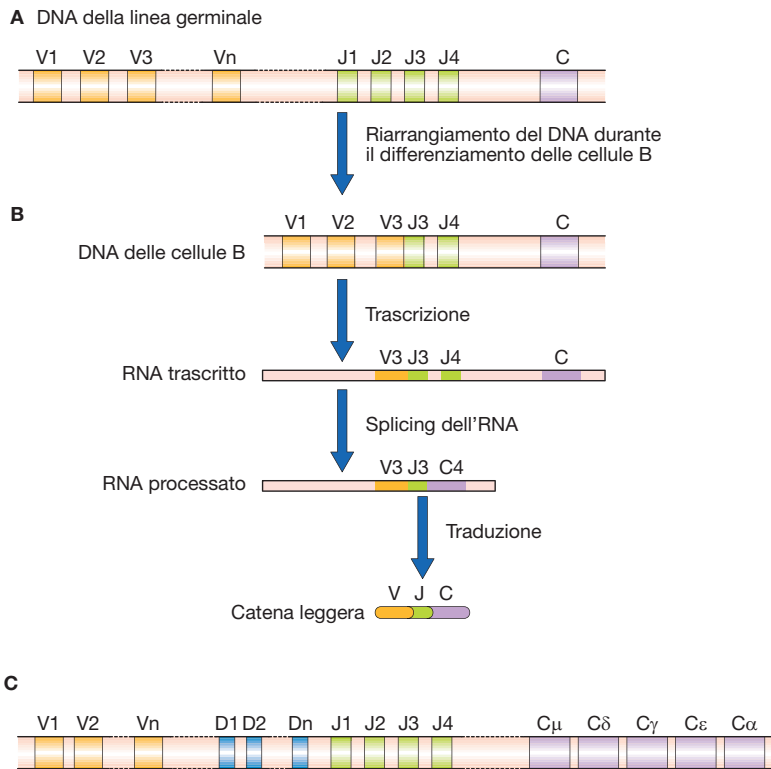


Figura 8.18 Meccanismo di ricombinazione V(D)J per generare una grande varietà di anticorpi. (A) Organizzazione strutturale del DNA codificante per una parte della catena leggera di un anticorpo nel contesto del genoma di una cellula in cui la ricombinazione V(D)J non è ancora avvenuta. (B) Risultato del riarrangiamento strutturale sul DNA determinato da un evento di ricombinazione tra la regione V3 e quella J3 durante il differenziamento di una cellula B. Quello indicato rappresenta solo uno degli eventi di ricombinazione possibili. Nella figura è anche mostrato che il gene in questione verrà trascritto dall'RNA polimerasi, processato dal meccanismo di splicing e l'mRNA maturo tradotto per generare la catena leggera indicata in figura. (C) È schematizzata la situazione originale presente in una cellula pre-B in una regione codificante per la catena pesante delle immunoglobuline: si notino la regione D aggiuntiva e i molteplici segmenti per le regioni costanti indicati con le lettere greche.

un locus specifico per la catena pesante di un anticorpo di topo è costituito da più di 100 segmenti V, 12 regioni D e 4 regioni J. La ricombinazione V(D)J che coinvolge tutti i possibili segmenti del gene può generare circa 5000 sequenze proteiche diverse. Poiché un anticorpo è costituito da una qualsiasi coppia di catene leggere e pesanti, la diversità di anticorpi che si possono generare diventa molto grande.

Senza entrare in dettagli che esulano dal nostro obiettivo, la ricombinazione V(D)J avviene con un meccanismo simile alla escissione dei trasposoni: infatti, i segmenti genici V(D)J che vengono riarrangiati sono fiancheggiati da coppie di sequenze segnale per la ricombinazione organizzate come ripetizioni invertite che sono analoghe a quelle che si trovano alle estremità dei trasposoni a DNA. La ricombinazione nei geni delle immunoglobuline è iniziata dai prodotti dei geni *Rag1* e *Rag2* (*Recombination Activating Gene*) che causano dei DSB specifici in tali regioni geniche. Sebbene la riparazione di DSB possa avvenire, come abbiamo già visto, sia tramite meccanismi di ricombinazione omologa che tramite meccanismi di *Non-Homologous End Joining* (NHEJ), la risoluzione dei DSB generati in modo RAG1/2-dipendente avviene quasi esclusivamente tramite NHEJ e l'azione delle proteine coinvolte è molto simile a quella di una trasposasi.

Alcuni degli esempi descritti in questo capitolo dovrebbero avere dato un'idea dell'importanza dei trasposoni in diversi aspetti della regolazione dei geni e della funzionalità dei cromosomi. L'aspetto, però, forse più rilevante dei trasposoni è la loro importanza nel determinare la plasticità del genoma. Il crossing over fra trasposoni presenti in siti cromosomici non omologhi può generare duplicazioni e inserzioni. Il genoma umano contiene, per esempio, un numero molto elevato di duplicazioni, e circa un terzo di esse sono fiancheggiate da elementi *Alu*, suggerendo che derivano da appaiamenti non corretti fra trasposoni *Alu* e successiva ricombinazione.

Bisogna anche ricordare che la capacità dei trasposoni di saltare all'interno di un gene è uno dei sistemi più utilizzati, soprattutto nelle piante ma anche nei batteri,

per generare l'inattivazione di uno specifico gene (knock-out). Infatti, l'inserimento del trasposone (mutagenesi inserzionale) causa un'inserzione di DNA esogeno nel gene determinandone l'inattivazione. La posizione del trasposone, e quindi del gene inattivato, è poi deducibile tramite PCR o tecniche di "etichettatura" o "gene tagging" del gene inattivato. Alcune delle tecniche più comuni per inattivare un gene o per spegnerne l'espressione sono descritte nel paragrafo 20.11.

Infine, in questo capitolo è necessario descrivere brevemente i retrovirus (**finestra 8.3**) che sono molto simili ai retrotrasposoni LTR anche se questi ultimi, a differenza dei retrovirus, non sono in grado di generare particelle virali infettive. I retrovirus sono tra i più importanti virus che infettano cellule eucariotiche e causano gravi patologie nell'uomo, quale l'AIDS (*Acquired Immuno Deficiency Syndrome*).

Lecture di approfondimento

- Chen, J.M., Cooper, D.N., Chuzhanova, N., Ferec, C. e Patrinos, G.P. (2007), «Gene conversion: mechanisms, evolution and human diseases», *Nature Reviews Genetics*, **8**, pp. 762-775.
- Haber, J.E. (1998), «Mating-type gene switching in *Saccharomyces cerevisiae*», *Annual Review of Genetics*, **32**, pp. 561-599.
- Haber, J.E., Ira, G., Malkova, A. e Sugawara, N. (2004), «Repairing a double-strand break by homologous recombination: revisiting the Robin Holliday's model», *Philosophical Transactions of the Royal Society London Biological Science*, **359**, pp. 79-86.
- Krogh, B.O. e Symington, L.S. (2004), «Recombination proteins in yeast», *Annual Review of Genetics*, **38**, pp. 233-271.
- Paques, F. e Haber, J.E. (1999), «Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*», *Microbiology and Molecular Biology Reviews*, **63**, pp. 349-404.
- San Filippo, J., Sung, P. e Klein, H. (2008), «Mechanism of eukaryotic homologous recombination», *Annual Review of Biochemistry*, **77**, pp. 229-257.
- Taylor, J.E. e Rudenko, G. (2006), «Switching trypanosome coats: what's in the wardrobe?», *Trends in Genetics*, **22**, pp. 614-620.

D

TRASCRIZIONE E SUE REGOLAZIONI

L'espressione dei geni ha, in tutti gli organismi viventi, un ruolo centrale nelle relazioni tra genotipo e fenotipo. Il flusso dell'informazione genetica dal DNA alle proteine passa attraverso un complesso processo cellulare, la trascrizione, che permette la sintesi regolata di un RNA messaggero che, a sua volta, trasferisce alla macchina molecolare del ribosoma le istruzioni necessarie per assemblare una proteina.

La trascrizione è finemente regolata nei batteri, nei virus e negli organismi eucarioti. Come vedremo in questa sezione, l'evoluzione ha portato a differenziare e rendere altamente specifica la regolazione della trascrizione, anche se alcuni meccanismi di base sono altamente conservati. L'RNA polimerasi batterica ha dei domini molto simili a quelli delle RNA polimerasi degli eucarioti, soprattutto nella regione responsabile della sintesi dell'RNA, mentre le regioni che devono interagire con proteine regolatrici sono diverse sia nei batteri che nelle tre polimerasi eucariotiche.

I principi di base della regolazione dei geni sono stati scoperti a metà degli anni sessanta, soprattutto grazie al lavoro di François Jacob e Jacques Monod, che dimostrarono che la regolazione è dovuta alle interazioni tra sequenze che agiscono in cis, quali promotori, operatori e altre sequenze di legame per fattori vari, e proteine che agiscono in trans, come le RNA polimerasi e vari tipi di fattori di trascrizione. I promotori sono riconosciuti dalle RNA polimerasi e contengono tutte le informazioni necessarie per un accurato inizio della trascrizione. Altre sequenze di DNA specifiche sono riconosciute da proteine che possono inibire la trascrizione agendo da repressori, o attivarla agendo come proteine attivatrici. Le funzioni degli attivatori e dei repressori, a loro volta, possono essere modulate a livello fisiologico, regolando l'espressione dei corrispondenti geni.

Mentre i principi di base della regolazione trascrizionale sono comuni a tutti gli organismi, alcuni aspetti di tale regolazione sono profondamente diversi tra procarioti ed eucarioti e questa diversità è legata principalmente al fatto che, mentre nei batteri il DNA è più disponibile al riconoscimento diretto da parte della polimerasi e delle proteine regolatrici, negli eucarioti il DNA è impaccato in una struttura cromatinica complessa che impone altre strategie per il riconoscimento e la funzione dei promotori.

9

Trascrizione nei procarioti

9.1 Unità di trascrizione

9.2 Le tre fasi della trascrizione

9.3 Struttura delle RNA polimerasi

9.4 Inizio della trascrizione nei procarioti

9.5 Distacco dell'RNA polimerasi dal promotore e allungamento dell'RNA

9.6 Terminazione della trascrizione nei procarioti

Il flusso dell'informazione genica va dal DNA all'RNA e, da questo, alle proteine. Il processo di sintesi dell'RNA su uno stampo di DNA è chiamato **trascrizione** e l'attore primario dell'intero meccanismo è un enzima chiamato **RNA polimerasi**. Questo enzima è una complessa macchina molecolare formata da varie subunità che dopo essersi legata al DNA lo apre, creando una zona denaturata, chiamata **bolla di trascrizione**, che fornisce all'enzima il filamento stampo da cui è diretta la sintesi dell'RNA. La bolla si muove con l'enzima e il DNA si apre via via che l'enzima procede e si richiude posteriormente.

Come illustrato in **figura 9.1** il processo di trascrizione comprende tre diverse fasi: inizio, allungamento e terminazione. L'RNA prodotto non rimane appaiato al DNA, ma si stacca dallo stampo a una distanza di pochi nucleotidi da dove è stato aggiunto l'ultimo ribonucleotide alla catena. Nei batteri, ciò permette all'RNA

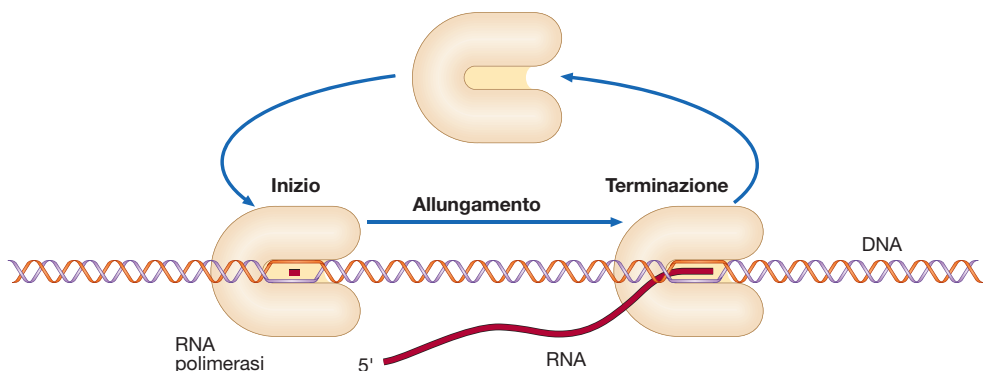


Figura 9.1 Le tre fasi della trascrizione. L'RNA polimerasi riconosce il promotore e inizia la trascrizione dal sito di inizio; estende poi il trascritto fino al sito di terminazione.

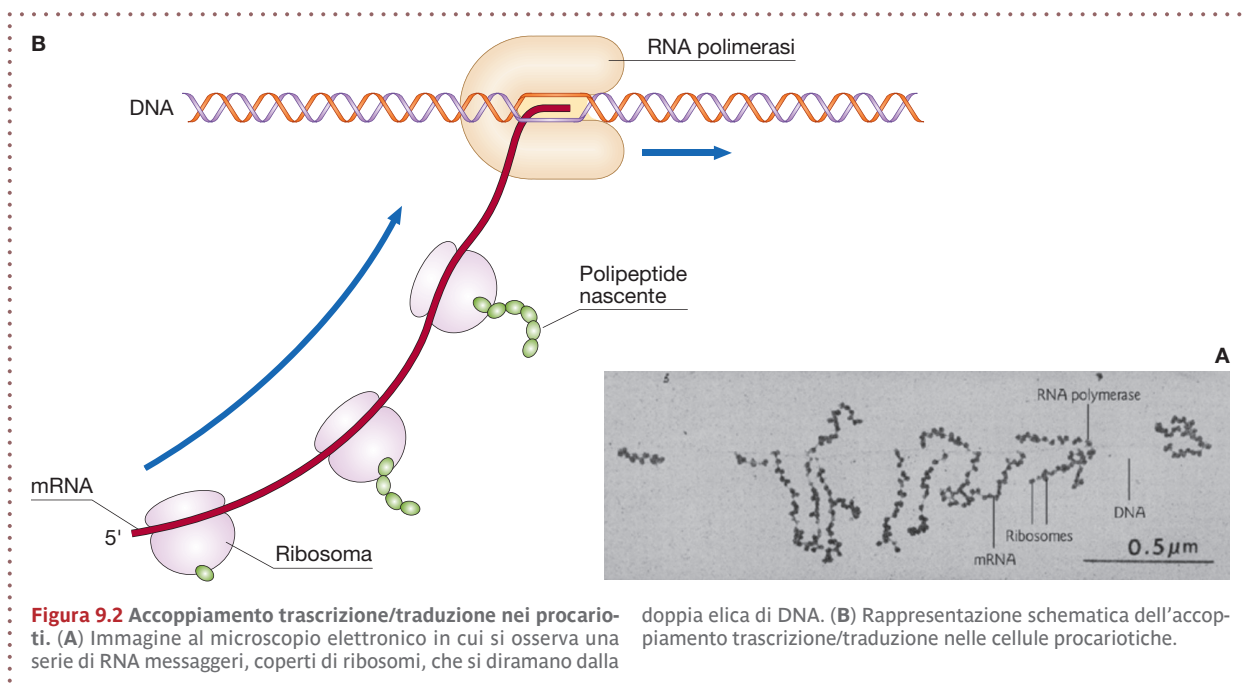


Figura 9.2 Accoppiamento trascrizione/traduzione nei procarioti. (A) Immagine al microscopio elettronico in cui si osserva una serie di RNA messengeri, coperti di ribosomi, che si diramano dalla

doppia elica di DNA. (B) Rappresentazione schematica dell'accoppiamento trascrizione/traduzione nelle cellule procariotiche.

di essere immediatamente disponibile per la sintesi di proteine, come è illustrato in **figura 9.2**, dove, nell'immagine al microscopio elettronico (**fig. 9.2A**), si può osservare una serie di RNA messengeri che diramano dalla catena di DNA e che sono legati da numerosi ribosomi; l'immagine suggerisce che più molecole di RNA polimerasi possono trascrivere lo stesso gene contemporaneamente, producendo più copie di RNA messaggero che vengono usate dai ribosomi per sintetizzare la corrispondente proteina. L'accoppiamento trascrizione/traduzione tipico della cellula procariotica è illustrato schematicamente in **figura 9.2B**.

9.1 Unità di trascrizione

Quando la doppia elica del DNA viene aperta, l'RNA polimerasi usa uno solo dei due filamenti come stampo (**fig. 9.3**). Riferendoci a geni codificanti per proteine, il **filamento stampo** è complementare al messaggero, mentre l'altro **filamento**, che viene chiamato **codificante**, ha la stessa sequenza del messaggero, con le T al posto delle U. L'RNA polimerasi sintetizza l'RNA in direzione $5' \rightarrow 3'$ e, quindi, quale dei due filamenti del DNA fungerà da stampo dipenderà dall'orientamento della polimerasi su quel tratto di DNA.

L'RNA polimerasi si lega a particolari sequenze sul DNA chiamate **promotori** e che sono situate all'inizio del gene. Il promotore contiene anche il nucleotide da cui inizia la sintesi dell'RNA, che viene definito **sito d'inizio della trascrizione (TSS)**. La trascrizione procede poi fino a una particolare sequenza chiamata **terminatore** e si definisce **unità di trascrizione** il tratto di DNA che va dal promotore fino al terminatore, espresso come una singola molecola di RNA (**fig. 9.4**). Nei batteri un'unità di trascrizione può comprendere anche più di un gene.

Le sequenze di DNA che precedono il punto di inizio della trascrizione sono spesso indicate, per convenzione, come sequenze "a monte" (*upstream*), mentre quelle che seguono l'inizio come sequenze "a valle" (*downstream*). La **figura 9.4** mostra anche come la posizione delle basi nella regione del promotore è numerata in entrambe le direzioni partendo dal punto d'inizio della trascrizione, che ha il valore di +1. I numeri a valle hanno valore positivo e quelli a monte valore negativo; il nucleotide che precede il punto d'inizio è identificato con -1.

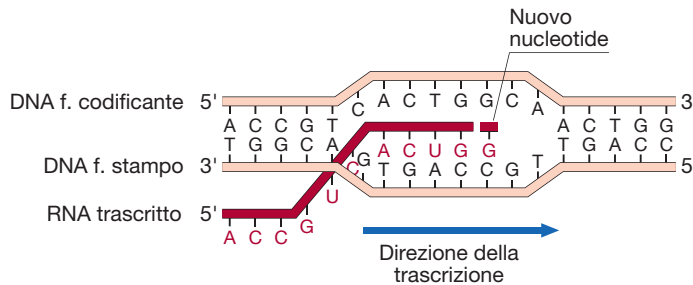


Figura 9.3 L'RNA polimerasi trascrive uno solo dei due filamenti della doppia elica del DNA. Nella trascrizione l'RNA polimerasi usa, come stampo, uno solo dei due filamenti di DNA. La sequenza dell'RNA trascritto è complementare a quella del filamento stampo ed equivalente a quella del filamento codificante.

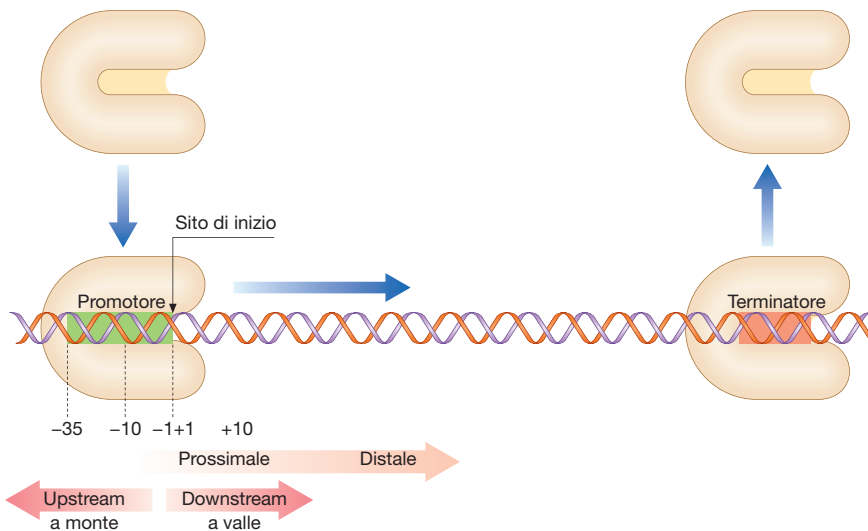


Figura 9.4 Un'unità di trascrizione è delimitata dal promotore e dal terminatore. Il punto di inizio della trascrizione divide la regione in sequenze "a monte" (*upstream*) in cui le basi sono numerate con valori negativi, e sequenze "a valle" (*downstream*) in cui le basi sono numerate con valori positivi. I termini "prossimale" e "distale" si riferiscono alla posizione delle regioni del gene rispetto al sito di inizio.

Come abbiamo accennato, la sintesi dell'RNA avviene all'interno di una bolla di DNA aperto che è formata dalla polimerasi dopo che si è legata al promotore (**fig. 9.5**). La sintesi della catena di RNA avviene in direzione 5'→3' appaiando i ribonucleotidi in maniera complementare sul filamento stampo. L'RNA polimerasi, a differenza della DNA polimerasi, non ha bisogno di un innesco per sintetizzare un nuovo filamento di RNA. Dopo il suo posizionamento, il gruppo 3'OH del primo nucleotide reagisce con il successivo nucleoside 5' trifosfato il cui fosfato in posizione α è usato per formare il legame fosfodiesterico, mentre i fosfati β e γ sono rilasciati come una molecola di pirofosfato.

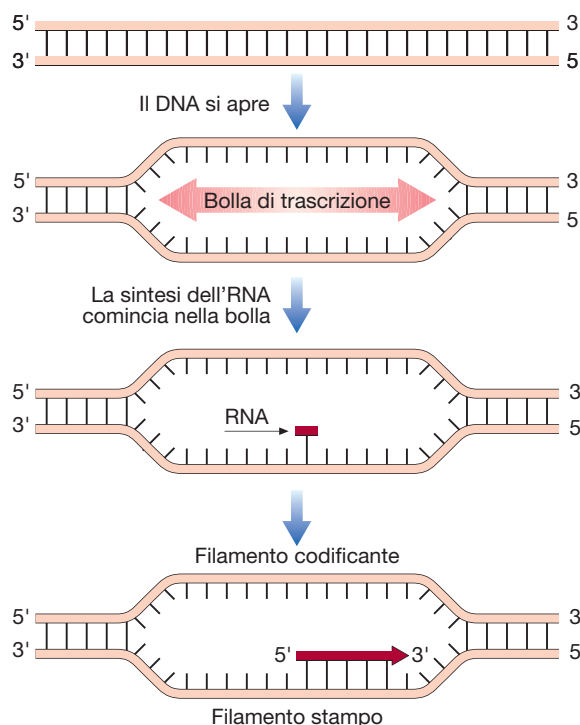
Nei batteri la velocità di sintesi è circa 50 nucleotidi al secondo a 37 °C che corrisponde bene alla velocità della sintesi proteica, che è di 15 amminoacidi al secondo, rendendo possibile il contemporaneo svolgersi dei due processi. Durante l'avanzamento lungo il DNA, l'RNA polimerasi controlla al suo interno l'apertura e la chiusura della bolla, rispettivamente a valle e a monte rispetto al suo senso di marcia.

La bolla di trascrizione è lunga circa 25 pb, ma il tratto che forma un ibrido tra DNA e RNA è lungo 8-9 pb; questa struttura è comune a tutte le RNA polimerasi sia batteriche che eucariotiche.

9.2 Le tre fasi della trascrizione

Come abbiamo accennato, l'RNA polimerasi per trascrivere un tratto di DNA che contiene uno o più geni procede con tre distinti passaggi (**fig. 9.6**): la **fase d'inizio**, in cui avviene il riconoscimento del promotore e il DNA si apre for-

Figura 9.5 Bolla di trascrizione e sintesi dell'RNA. Nella fase di inizio della trascrizione l'RNA polimerasi si lega al promotore e provoca l'apertura della doppia elica formando così la "bolla di trascrizione". Senza bisogno di un "innesco", il primo nucleoside trifosfato si posiziona esponendo il 3'-OH che reagisce con il successivo dando così inizio alla trascrizione.



mando la bolla di trascrizione per iniziare la sintesi dell'RNA; la **fase d'allungamento**, in cui la polimerasi e la bolla si muovono lungo il DNA estendendo la catena di RNA; la **fase di terminazione**, in cui l'RNA polimerasi si arresta al terminatore, il trascritto di RNA si dissocia dalla polimerasi e la bolla si richiude.

L'inizio è la fase più complessa del processo perché l'RNA polimerasi deve riconoscere il promotore in maniera specifica e legarsi stabilmente ad esso: vedremo successivamente in dettaglio come avviene questo riconoscimento. Il legame che si forma inizialmente tra l'RNA polimerasi e il promotore viene definito **complesso chiuso** perché il DNA non è stato ancora aperto per formare la bolla (fig. 9.6). Una volta che l'enzima è stabilmente legato al promotore, una serie di cambiamenti conformazionali al suo interno promuove l'apertura della bolla e la formazione del **complesso aperto**. Questa apertura rende disponibile il filamento stampo al riconoscimento complementare dei nucleotidi che entrano nel complesso e la sintesi può iniziare. Inizialmente l'RNA polimerasi sintetizza, senza distaccarsi dal promotore, un frammento di RNA che viene rilasciato prima di raggiungere i 9 nt di lunghezza. Questa sintesi abortiva è ripetuta più volte fino a quando, superata la lunghezza di 9 nt, si forma un ibrido DNA-RNA sufficientemente stabile. La polimerasi può a questo punto rilasciare il promotore e proseguire nella fase di allungamento. Studi recenti hanno dimostrato che il DNA, entrando nel complesso, fa da regolatore allosterico, modificando la struttura stessa della polimerasi.

Durante la fase di allungamento l'enzima si muove lungo il filamento stampo in direzione $3' \rightarrow 5'$ sintetizzando la catena di RNA dal $5'$ al $3'$ e muove con sé la bolla di trascrizione; il DNA si apre nella direzione della sintesi e si richiude alle sue spalle, mantenendo la bolla di una lunghezza costante. Durante questa fase l'enzima, che come abbiamo detto si muove in media a una velocità di circa 50 nucleotidi al secondo, può andare incontro a rallentamenti che dipendono dalle sequenze di DNA che incontra lungo il suo cammino. Alcune sequenze sono in grado di rallentare la polimerasi fino a fermarla e questo può provocare una terminazione precoce. In questa fase, se viene incorporato un nucleotide non corretto, l'appaiamento di

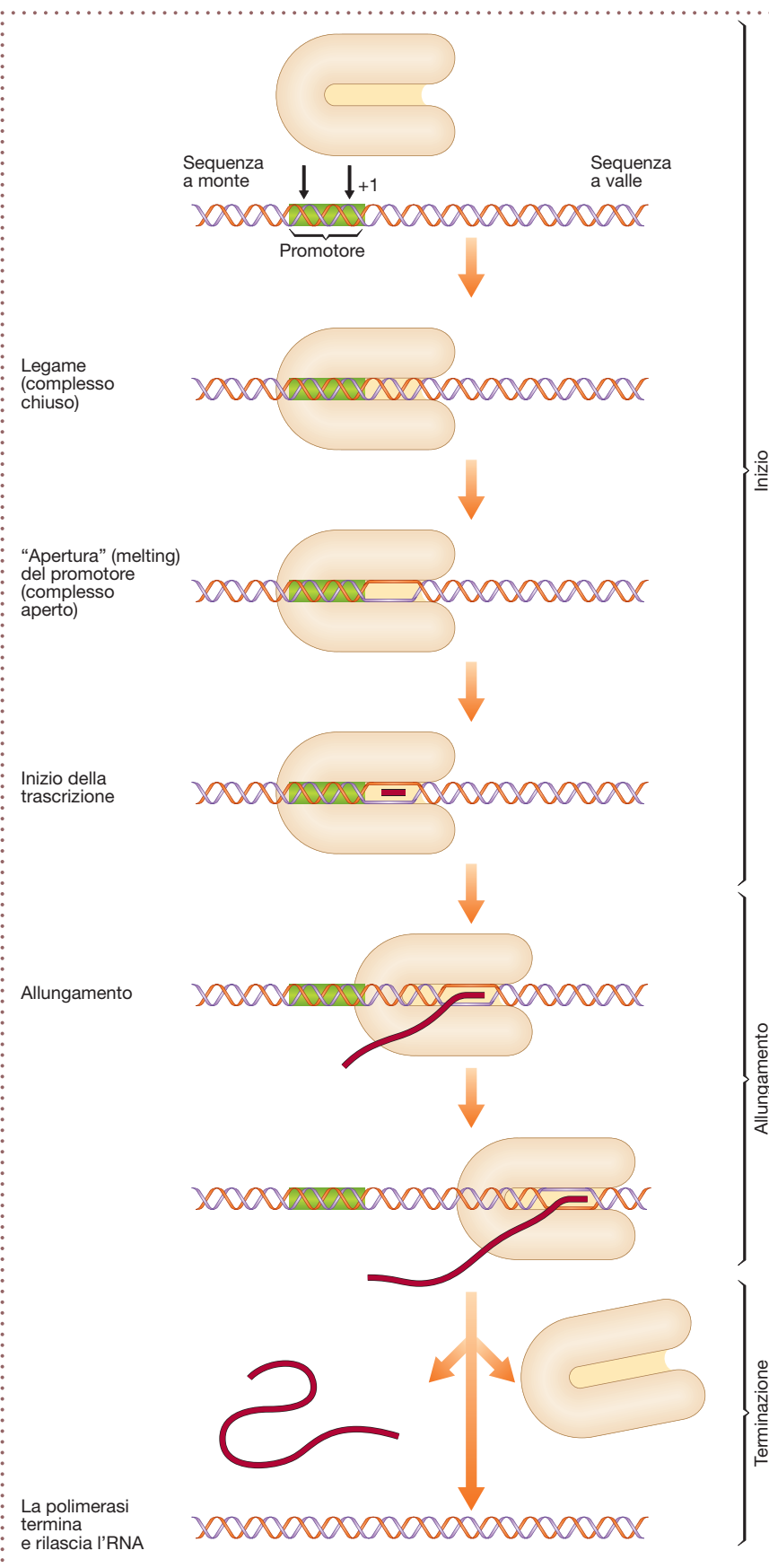


Figura 9.6 Le tre fasi della trascrizione. La polimerasi si lega in modo specifico al promotore e si forma il complesso chiuso. Successivamente si forma il complesso aperto e, dopo una fase abortiva, inizia l'allungamento del trascritto. La polimerasi raggiunge il terminatore e si stacca dal DNA.

basi nella bolla risulterà anomalo e questo indurrà la polimerasi a tornare indietro e a eliminare il nucleotide sbagliato con un meccanismo di correzione.

La terminazione è l'ultima fase della trascrizione in cui l'RNA polimerasi rilascia il filamento di RNA prodotto e si dissocia dal DNA. Nei batteri esistono delle sequenze specifiche alla fine di ogni gene, i terminatori, che permettono il distacco della polimerasi e che analizzeremo in dettaglio più avanti.

9.3 Struttura delle RNA polimerasi



Figura 9.7 Seth A. Darst. I grandi progressi delle conoscenze del macchinario e del meccanismo della trascrizione sono il risultato del lavoro di numerosi gruppi di ricerca. Tra questi, Seth Darst ha dato un contributo essenziale risolvendo la struttura tridimensionale dell'enzima di *Thermus aquaticus*.

La nostra conoscenza dei meccanismi molecolari del funzionamento dell'RNA polimerasi è dovuta al lavoro di numerosi gruppi di ricerca, tra cui quello di Seth A. Darst (**fig. 9.7**) e collaboratori, che hanno risolto la struttura tridimensionale sia dell'enzima batterico che degli enzimi eucariotici. La **figura 9.8** mostra le strutture dell'RNA polimerasi procariotica di *Thermus aquaticus* e dell'RNA polimerasi II eucariotica del lievito. In ambedue i casi l'enzima ha una forma caratteristica a "pinza" che gli permette di agganciarsi al DNA. Esso è costituito da diverse subunità che sono descritte schematicamente nella parte inferiore della figura, dove i colori mettono in evidenza la similitudine dei vari domini nei due enzimi. Come si vede, la struttura dell'enzima è molto conservata e soprattutto la parte centrale (o *core* = nucleo dell'enzima), dove avviene la sintesi, è davvero molto simile.

Mentre nei batteri è presente una sola RNA polimerasi che è responsabile di tutta la sintesi dell'RNA che avviene nella cellula, nei nuclei delle cellule eucariotiche esistono tre diversi enzimi che trascrivono diverse classi di geni (delle polimerasi eucariotiche parleremo a fondo nel par. 11.1). Va anche detto che molti studi sul meccanismo della trascrizione si sono avvalsi anche di RNA polimerasi più semplici, come l'RNA polimerasi del batteriofago T7, la cui struttura è mostrata in **figura 9.9**. Qui si vede come anche questo enzima, che è costituito da una singola subunità di 99 kDa, presenta una forma a pinza. Esso è in grado di legarsi da solo e con molta efficienza al suo promotore del fago T7, senza l'intervento di alcun fattore accessorio.

Figura 9.8 Strutture della polimerasi batterica e della polimerasi II degli eucarioti. (A) Strutture cristallografiche del nucleo (*core*) dell'RNA polimerasi del batterio *Thermus aquaticus* e dell'RNA polimerasi II del lievito. In ciascuno dei due enzimi le varie subunità sono rappresentate con colori diversi, mentre le subunità omologhe nei due enzimi hanno lo stesso colore. (B) Rappresentazione schematica delle due RNA polimerasi che, utilizzando lo stesso codice colori, permette di identificare le varie subunità mostrate in (A).

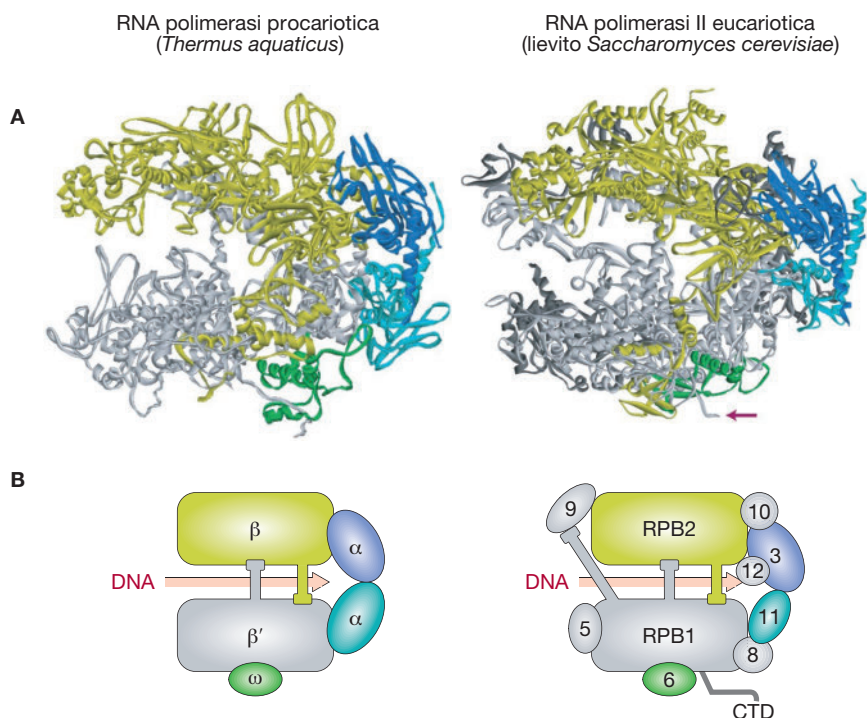




Figura 9.9 RNA polimerasi del fago T7. La più semplice RNA polimerasi è quella del fago T7: è stata utilizzata come modello in molti studi sul meccanismo della trascrizione. È costituita da una singola subunità ed è in grado di legarsi da sola con grande specificità al suo promotore. [Fonte: Jeruzalmi, D., Steitz, T.A. (1998), *EMBO J.*, 17, pp. 4101-4113.]

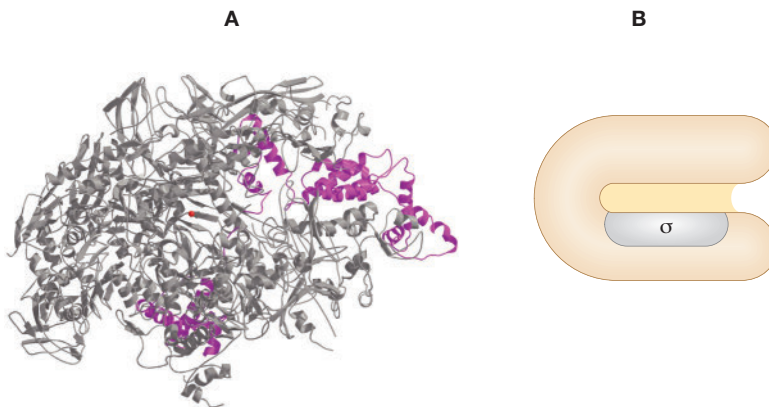


Figura 9.10 Struttura tridimensionale dell'RNA polimerasi batterica legata al fattore σ . (A) Struttura cristallografica dell'RNA polimerasi oloenzima di *Thermus aquaticus*: in viola si possono vedere i domini che costituiscono il fattore σ . [Fonte: Murakami, K.S. et al. (2002), *Science*, 296(5571), pp. 1285-1290.] (B) Rappresentazione schematica che useremo in questo testo per rappresentare l'RNA polimerasi oloenzima procariotica.

Vediamo ora in dettaglio come è fatta e come funziona l'RNA polimerasi batterica. L'enzima misura nel suo lato più lungo circa 160 Å e negli altri due lati 90 e 95 Å, rispettivamente. Il peso molecolare del nucleo (core) dell'enzima è di circa 400 kDa e, come si vede tornando alla **figura 9.8**, è costituito di 5 subunità: 2 subunità α , una subunità β , una β' e una subunità ω . Le due subunità α sono responsabili dell'assemblaggio del complesso e contengono due domini che hanno funzioni diverse: il dominio C-terminale (chiamato α -CTD, da non confondere con il CTD dell'RNA polimerasi II eucariotica che vedremo più avanti) si lega a una zona del promotore chiamata *UP-element* (elemento a monte), mentre il dominio N-terminale è responsabile dell'interazione con le altre subunità dell'enzima. La subunità β contiene il sito catalitico, che sintetizza l'RNA, di cui fanno parte anche due atomi di Mg^{2+} che sono essenziali per la sintesi. Uno di questi atomi è sempre presente nel sito attivo, l'altro viene trasportato nel complesso dai nucleotidi in entrata. La subunità β' si lega al DNA in maniera non specifica e, infine, la subunità ω ha la funzione di promuovere e mantenere stabile il complesso con un meccanismo simile ai chaperoni molecolari (*vedi* par. 19.1).

Per legarsi in maniera specifica al promotore, la polimerasi ha bisogno di una sesta subunità, chiamata σ . La **figura 9.10** mostra la struttura cristallografica dell'enzima completo $\alpha_2\beta\beta'\omega\sigma$ (od oloenzima), che ha un peso molecolare di circa 480 kDa. Nella figura si può vedere che il fattore σ è costituito da quattro domini, o regioni, distribuiti sul nucleo enzimatico, in parte verso l'esterno per riconoscere e legare il promotore e in parte nella regione interna tra le due parti della "pinza", formata dalle subunità β' e β , occupando parzialmente il canale dove si posiziona il DNA. Nella figura è anche mostrata la rappresentazione

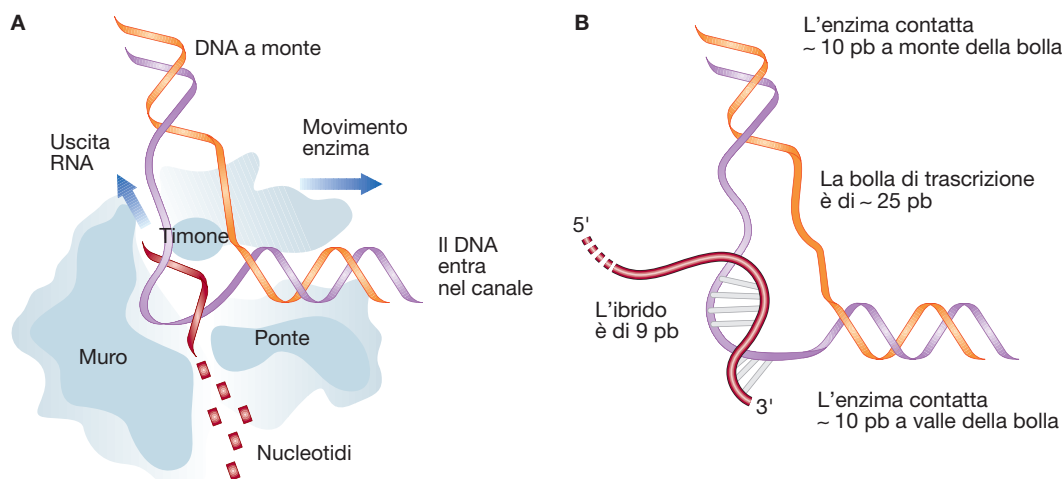


Figura 9.11 Principali domini strutturali interni della RNA polimerasi. (A) Rappresentazione schematica dei domini strutturali della RNA polimerasi e della loro relazione con il DNA nel complesso aperto. (B) La bolla di trascrizione con la regione ibrida RNA-DNA di 9 nucleotidi.

so aperto. (B) La bolla di trascrizione con la regione ibrida RNA-DNA di 9 nucleotidi.

schematica che usiamo in questo testo. Il fattore σ funziona come un vero e proprio fattore di trascrizione. Infatti, con σ legato si riduce la capacità globale della polimerasi di legarsi al DNA, ma aumenta moltissimo la sua affinità per il promotore. Questa caratteristica permette all'RNA polimerasi, in assenza del fattore σ , di essere generalmente sempre legata al DNA e di non disperdersi nella cellula ma, con σ legato, di essere in grado di riconoscere con grande precisione i promotori. A seconda delle condizioni di crescita della cellula, dalle 3000 alle 5000 molecole di RNA polimerasi, delle circa 7000 che si trovano in una cellula di *Escherichia coli*, sono normalmente impegnate attivamente nella trascrizione, mentre le altre sono legate al DNA in modo non-specifico.

Al suo interno l'RNA polimerasi ha un solco lungo circa 55 Å con una larghezza di circa 25 Å. Esso può accogliere un tratto di circa 15-16 pb di doppia elica di DNA che ha un diametro di 20 Å. La **figura 9.11** illustra le varie strutture interne dell'enzima e il ruolo che hanno nella formazione del complesso con il DNA e nella sintesi dell'RNA. Come si può vedere, il DNA è costretto dalla struttura a fare una piegatura di quasi 90° nella parte posteriore dell'enzima che viene definita "muro". Nella parte superiore c'è il foro di uscita dell'RNA e più a destra il canale da dove esce il DNA, che si riappaia alla fine della bolla; al centro c'è una struttura, chiamata "timone", che contribuisce a tenere aperta la bolla. Nella parte inferiore c'è una struttura a "imbuto" che permette l'entrata dei ribonucleosidi trifosfati. Nella parte destra della figura si vede in dettaglio la bolla di trascrizione e la zona ibrida DNA-RNA di 9 nucleotidi.

9.4 Inizio della trascrizione nei procarioti

Nella fase di inizio l'RNA polimerasi non ha bisogno di un primer, ma è in grado di iniziare direttamente la sintesi dell'RNA sullo stampo. In genere l'enzima inizia quasi sempre i trascritti con una A e questo suggerisce che nel sito attivo vi sia una particolare affinità per il corrispondente nucleoside trifosfato che gli permette di avere un legame abbastanza stabile da poter essere poi legato al nucleotide successivo. Entrambi devono essere posizionati correttamente per appaiarsi con lo stampo e realizzare interazioni di impilamento tra loro. Nel caso dei batteri si pensa che anche una regione del fattore σ , chiamata **loop σ_2/σ_3** (vedi oltre), prenda parte a questo riconoscimento, dal momento che mutazioni

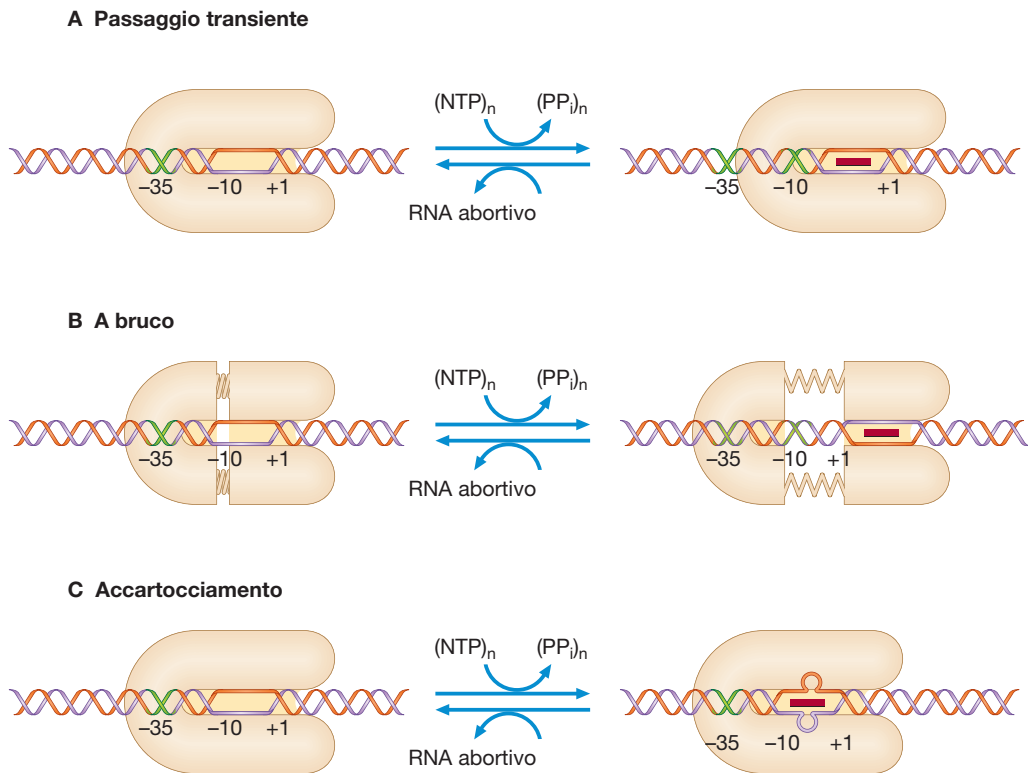


Figura 9.12 I tre modelli che descrivono la fase iniziale della sintesi dell'RNA. (A) La polimerasi si sposta avanti e indietro sul promotore. (B) La polimerasi, che deve contenere una regione estensibile, si allunga sul DNA occupando una zona più vasta del suo ingombro in assenza dei nucleotidi. (C) La polimerasi richiama al suo interno un tratto di DNA "downstream" che si apre e si accartocchia al suo interno. Questo accartocciamento si pensa fornisca energia per promuovere il distacco dal promotore.

in questa regione influenzano la concentrazione di nucleotidi necessaria all'inizio della sintesi dell'RNA.

Come già detto, nella fase iniziale della sintesi l'RNA polimerasi sintetizza e poi rilascia dei corti frammenti di RNA, fino al raggiungimento di una lunghezza di 9-12 nt, che permette all'RNA di spazzare il loop σ_2/σ_3 , che occupa il canale di uscita. Questa fase iniziale abortiva è stata molto importante per comprendere come la polimerasi avanza sul DNA. Questi studi sono stati fatti usando diverse tecniche sperimentali, tra le quali il **footprint** (vedi par. 20.11). Questa tecnica ha permesso di vedere che lo spazio occupato dalla polimerasi sul promotore in presenza di nucleotidi (trascrizione attiva) è maggiore che in loro assenza (trascrizione bloccata). Per dare una spiegazione a questi dati sono stati proposti tre possibili modelli (**fig. 9.12**). In un primo modello, detto **passaggio transiente**, la polimerasi avanza per un breve tratto sintetizzando un corto frammento abortivo di RNA e poi indietreggia ritornando al punto di partenza sul promotore. In un secondo modello, chiamato **a bruco**, proprio come questo animale la polimerasi, che secondo questa ipotesi deve avere una struttura molto flessibile, si allunga e si contrae sul DNA. Nella terza ipotesi chiamata **accartocciamento**, in inglese *scrunch*, la polimerasi inizialmente sta ferma sul promotore e ingloba dentro di sé un tratto di DNA che si "accartocchia" al suo interno. È possibile che la polimerasi proceda anche nella fase di allungamento usando questo sistema. Quest'ultima ipotesi è sostenuta da recenti dati di cristallografia, che hanno permesso di vedere all'interno della struttura un tratto più lungo di DNA, ma soprattutto da una moderna tecnica di analisi chiamata **FRET** (vedi par. 20.11 e fig. 20.63). Questa tecnica permette di marcare con atomi fluore-

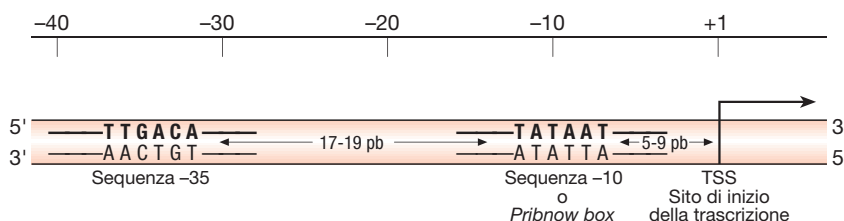
scenti dei nucleotidi sul DNA e successivamente calcolarne la relativa distanza. Si è visto quindi che due zone del DNA all'interno dell'enzima vengono a trovarsi vicine dopo essere state "accartocciate" dalla polimerasi. Anche nella fase di distacco dal promotore l'accartocciamento è importante e le ipotesi più recenti sostengono che lo svolgimento di un giro completo dell'elica del DNA in entrata e il suo accartocciamento all'interno della polimerasi forniscano l'energia necessaria a rompere le interazioni dell'oloenzima con il promotore.

Promotori e fattore sigma

Nel paragrafo precedente abbiamo visto come funziona l'RNA polimerasi durante la fase di sintesi dell'RNA e, inoltre, abbiamo detto che di per sé questo enzima ha un'affinità intrinseca al DNA; quindi, nella cellula la maggior parte delle RNA polimerasi si trova legata al DNA in un equilibrio in cui il legame sarà veloce e il distacco lento. Questo legame al DNA non è specifico, mentre per trascrivere un determinato gene l'enzima deve essere in grado di riconoscerne il sito di inizio con grande precisione. Questo problema è risolto da due diversi elementi: la sequenza specifica del DNA chiamata **promotore** e la **subunità σ** che si associa al core della polimerasi formando l'oloenzima $\alpha_2\beta\beta'\omega\sigma$. Il fattore σ riconosce il promotore, costituito da sequenze conservate che si trovano a monte del sito di inizio della trascrizione, ma solo quando è parte integrante dell'RNA polimerasi oloenzima. Infatti σ da solo non è in grado di riconoscere e legare il promotore. Esistono, come vedremo più avanti, diversi fattori σ ; in *E. coli* il fattore σ più usato è σ^{70} , chiamato così perché ha un peso molecolare di 70 kDa, che riconosce un promotore costituito da due sequenze conservate, lunghe sei nucleotidi, che si trovano in posizione -10 e -35 rispettivamente. Come mostrato in **figura 9.13**, queste sequenze sono separate da un tratto lungo da 17 a 19 nucleotidi. Comparando più di 300 promotori presenti nel genoma di *E. coli* che vengono riconosciuti dal fattore σ^{70} , si è potuto generare una sequenza consenso che rappresenta gli elementi -10 e -35 più comuni, separati dalla sequenza intermedia più frequente di 17 nucleotidi. La sequenza consenso dell'elemento -10, chiamata anche *Pribnow box* dal nome del ricercatore che l'ha scoperta, è, sul filamento codificante, $T_{80}A_{95}T_{45}A_{60}A_{50}T_{95}$, dove i numeri indicano, in percentuale, la frequenza con la quale quella base è presente. L'elemento -35 ha come consenso la sequenza $T_{82}T_{84}G_{78}A_{65}C_{54}A_{45}$ (vedi anche finestra 21.1). Come si può immaginare, il fatto che questi elementi siano ricchi in A e T, li rende facilmente denaturabili per formare il complesso aperto. Non tutti i promotori contengono esattamente la sequenza consenso, ma esistono piccole variazioni che determinano la "forza" di un promotore; infatti un promotore con sequenza identica o quasi identica al consenso sarà molto "forte", mentre sarà progressivamente più "debole" quanto più se ne allontanerà.

Esistono poi, per σ^{70} , alcuni promotori la cui struttura devia da quella tipica che abbiamo descritto. In **figura 9.14** sono descritti tre diversi tipi di promotori presenti nel genoma di *E. coli*:

Figura 9.13 Struttura del promotore classico per σ^{70} di *Escherichia coli*. Questo tipo di promotore è costituito da tre elementi: le due sequenze consenso -35 e -10 (chiamata anche *Pribnow box*), e il sito di inizio della trascrizione (TSS = *Transcription Start Site*).



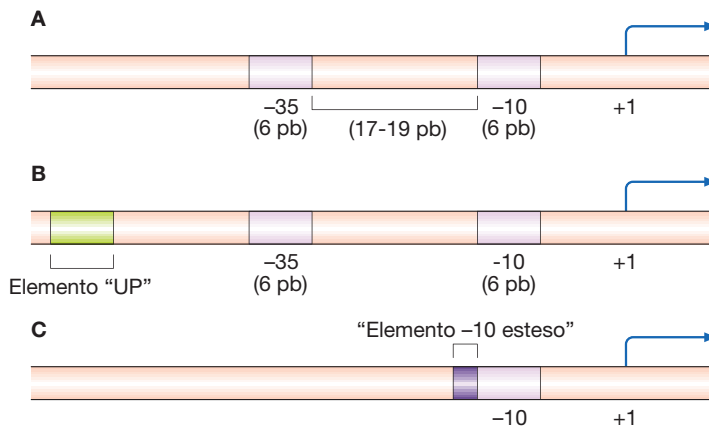


Figura 9.14 Confronto delle strutture dei tre tipi di promotori batterici per σ^{70} di *E. coli*. (A) Il promotore classico per σ^{70} . (B) Promotore contenente l'elemento aggiuntivo UP a monte della sequenza -35. (C) Promotore per σ^{70} che non contiene la sequenza -35, ma ha un elemento -10 esteso con una piccola sequenza conservata a monte del -10.

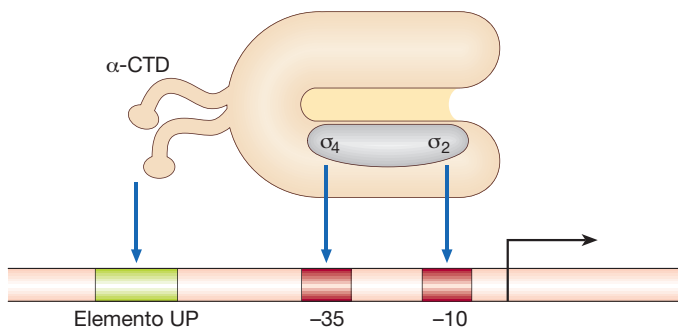


Figura 9.15 Interazione della RNA polimerasi con i promotori che contengono l'elemento UP. L'interazione non avviene solo tramite il fattore σ , ma anche con l' α -CTD (dominio C-terminale della subunità α dell'RNA polimerasi). Le regioni di σ di legame al DNA si trovano esposte in modo da permettere l'interazione e lo spazio di 75 Å che divide le regioni 2 e 4 corrisponde alla distanza tra le regioni -10 e -35 del promotore.

1. il promotore classico per σ^{70} ;
2. un promotore, implicato nella trascrizione dei geni per l'RNA ribosomiale, che contiene un elemento aggiuntivo a monte della sequenza -35, chiamato elemento UP; questa sequenza è riconosciuta dall' α -CTD (coda C-terminale della subunità α) dell'RNA polimerasi (vedi oltre). Quindi questi promotori sono riconosciuti e legati dall'RNA polimerasi sia tramite il fattore σ che tramite l' α -CTD della subunità α (fig. 9.15). Questo dominio è connesso al resto della proteina da una regione proteica definita "giunto flessibile" che permette di raggiungere l'elemento UP anche se non si trova immediatamente vicino alla regione -35;
3. un altro gruppo di promotori per σ^{70} , come per esempio quelli dei geni *gal*, che non contengono la sequenza -35 ma, in compenso, hanno un elemento -10 esteso contenente una piccola sequenza conservata a monte del -10.

Come abbiamo già visto in figura 9.10, il fattore σ è costituito da quattro regioni, o domini, che sono chiamate rispettivamente σ_1 , σ_2 , σ_3 e σ_4 . Un approfondimento sulle interazioni dell'RNA polimerasi con il fattore σ e i suoi domini è riportato nella **finestra 9.1**.

Diversi fattori sigma controllano promotori diversi

σ^{70} è il principale fattore che attiva la trascrizione promuovendo l'interazione della polimerasi oloenzima con il promotore; tuttavia, esistono altri fattori σ , alcuni dei quali presentati in **tabella 9.1**. Ognuno di essi riconosce promotori differenti che hanno sequenze -10 e -35 specifiche e diverse dalla sequenza consenso di σ^{70} .

FINESTRA 9.1

INTERAZIONI DEL FATTORE σ E DEI SUOI DOMINI CON L'RNA POLIMERASI E CON IL PROMOTORE

La risoluzione della struttura tridimensionale dell'RNA polimerasi con il fattore σ da parte del gruppo di Seth Darst della Rockefeller University ha permesso di analizzare a livello molecolare la **formazione del complesso aperto**. Abbiamo già detto nel testo e mostrato in figura 9.9 che il fattore σ è legato all'RNA polimerasi nella parte interna della "pinza" formata dalle subunità β e β' . Solo quando è assemblato con il nucleo (core) dell'RNA polimerasi a formare l'oloenzima, σ è in grado di riconoscere e legare specificamente il promotore. Il fattore σ è diviso in quattro regioni, o domini, chiamate rispettivamente σ_1 , σ_2 , σ_3 e σ_4 . Come illustrato schematicamente in **figura 9F1.1** la regione σ_2 riconosce l'elemento -10 , mentre la regione σ_4 riconosce l'elemento -35 . σ_4 è costituita da due α -eliche

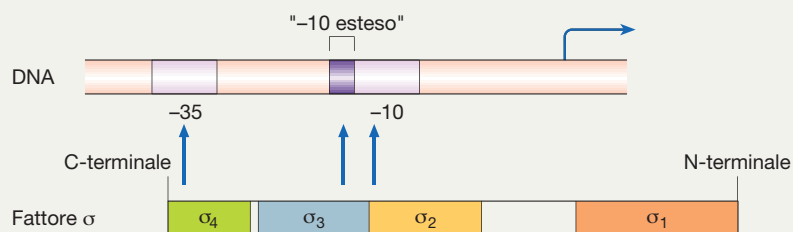


Figura 9F1.1 Interazioni tra le regioni del fattore σ e gli elementi che costituiscono i promotori batterici. Le regioni σ_2 e σ_4 interagiscono rispettivamente con le sequenze -10 e -35 ; la regione σ_3 interagisce con l'estensione della sequenza -10 quando questa è presente.

che formano un dominio elica-giro-elica (*helix-turn-helix*) che è tipico di molti fattori che si legano al DNA, come per esempio molti repressori batterici (*vedi par. 11.7*). In particolare, un'elica interagisce specificamente con le basi esposte nel solco maggiore della doppia elica e l'altra si posiziona sopra al solco interagen-

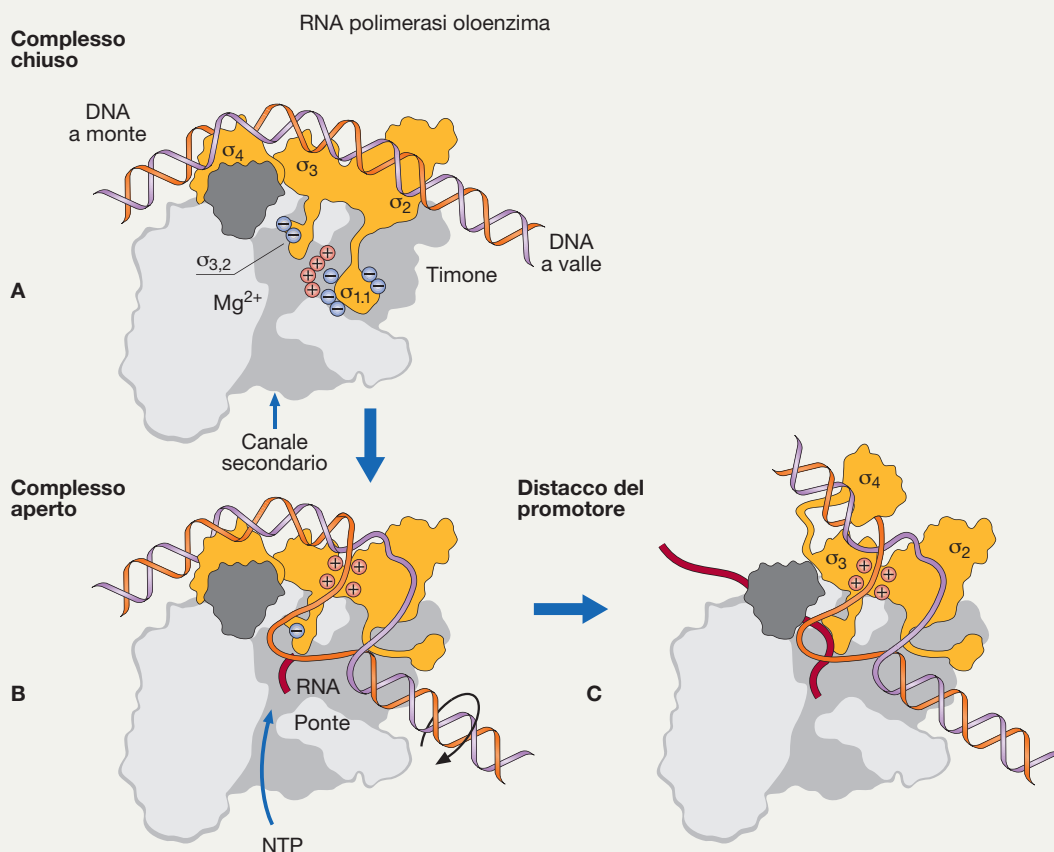


Figura 9F1.2 Struttura dell'oloenzima. (A) Il fattore σ , dopo l'interazione specifica con le sequenze -10 e -35 del promotore, va incontro a cambiamenti di conformazione dovuti al legame del DNA nel solco dell'RNA polimerasi. (B) Il cambiamento delle cariche intorno al dominio σ_1 ne promuove lo spostamento dal solco portando alla formazione del complesso aperto, dove avviene la

prima sintesi abortiva di brevi tratti di RNA. (C) Un secondo cambiamento conformazionale di σ_4 libera il foro di uscita dell'RNA e quindi promuove il distacco dal promotore e l'allungamento dell'RNA. L'RNA nascente è indicato in rosso; è anche indicato il canale secondario per l'ingresso dei nucleotidi (NTP) e le posizioni del Mg^{2+} e delle cariche positive e negative (+ e -).

do con la scheletro zucchero-fosfato. L'interazione di σ_2 con la regione -10 avviene sempre attraverso un' α -elica ma, in questo caso, i residui aromatici degli amminoacidi esposti sulla stessa faccia dell' α -elica interagiscono con le basi da -12 a -10, facendo un contatto con il filamento codificante del DNA. Il dominio σ_2 , infatti, è responsabile della transizione tra il complesso chiuso e il complesso aperto. L'interazione forte di σ_4 con la regione -35 è ancora specificamente l'oloenzima al DNA, mentre la regione σ_2 ha il compito di promuovere l'apertura del DNA. Nei promotori che contengono l'elemento -10 esteso, il contatto avviene con un' α -elica della regione σ_3 . La regione N-terminale di σ ha un ruolo di regolazione molto importante: infatti, se si elimina questo dominio, la proteina tronca è in grado di legarsi specificamente ai promotori anche in assenza di RNA polimerasi "core". Questo dato suggerisce che la regione N-terminale maschera le regioni σ_2 e σ_4 quando il fattore non è legato all'enzima "core" e quando non è in grado di occupare promotori in assenza delle subunità responsabili dell'attività di sintesi dell'RNA. Però, una volta che si forma l'oloenzima, un cambiamento conformazionale espone queste regioni rendendole in grado di legarsi al DNA.

La **figura 9F1.2** mostra uno spaccato della struttura della RNA polimerasi oloenzima dove si vedono i cambiamenti strutturali che avvengono tra la formazione del complesso chiuso e quella del complesso aperto. Come si può vedere in (A), σ_2 e σ_4 sono esposti all'esterno, contattando le regioni -10 e -35, mentre l'interno del canale di entrata del DNA è occupato dalla regione σ_1 , che ha carica negativa, come il DNA. Nella parte inferiore si vede il canale di entrata degli NTP. Dopo la formazione del complesso chiuso, cambiamenti conformazionali inducono l'apertura del DNA, per rendere disponibile il

filamento stampo, nella regione che va da -11 a +3. Questa modificazione strutturale non richiede, per σ^{70} , energia e viene definita **isomerizzazione** perché il complesso che si forma viene a trovarsi in una condizione energetica più favorevole. Un contributo a questa reazione può essere dato *in vivo* dal livello di torsione del DNA. Una volta formato il complesso aperto, la reazione è irreversibile, mentre il complesso chiuso può ancora dissociarsi dal DNA. Nella parte superiore della struttura è visibile il canale di uscita dell'RNA, che in questa fase è occupato dalle subunità σ_3 , σ_2 e σ_4 . Nella fase di isomerizzazione il DNA entra nel canale formato dalle due pinze β' e β calzando via la regione σ_1 , che si sposta di circa 50 Å venendosi a trovare praticamente all'esterno dell'enzima. Il DNA si apre cominciando dal nucleotide +3 e il filamento codificante, in verde chiaro, si sposta dal canale centrale, mentre il filamento stampo aderisce completamente alla regione del sito attivo, dove sono presenti i due atomi di Mg^{2+} . A questo punto inizia la sintesi e la polimerasi produce dei piccoli trascritti abortivi che vengono eliminati, fino a quando la reazione diventa stabile e il trascritto si allunga formando un ibrido DNA-RNA di 9 nucleotidi, all'interno della bolla. Il canale di uscita dell'RNA occupato da σ_3 e σ_4 viene liberato e σ_4 si stacca parzialmente dalla regione -35, permettendo in questo modo all'RNA sintetizzato di uscire dal canale. Questi distacchi progressivi di σ dal promotore promuovono la fase continua di allungamento. Il destino di σ è quello di staccarsi dal core enzimatico e impegnarsi nel legame con un'altra RNA polimerasi. Tuttavia, dati recenti suggeriscono che σ può rimanere legato al nucleo della RNA polimerasi, poiché mutazioni che influenzano il legame della regione σ_4 con la subunità β non alterano solo l'inizio della trascrizione, ma anche l'allungamento.

Tabella 9.1 Alcuni fattori σ alternativi in *Escherichia coli*.

Gene	Fattore σ	Funzione	Sequenza -35	Distanza	Sequenza -10
<i>rpoD</i>	σ^{70}	Generale	TTGACA	16-18 pb	TATAAT
<i>rpoH</i>	σ^{32}	Shock termico	CCCTTGAA	13-15 pb	CCCGATNT
<i>rpoN</i>	σ^{54}	Carenza di azoto	CTGGNA	6 pb	TTGCA
<i>fliA</i>	σ^{28}	Sistema del flagello	CTAAA	15 pb	GCCGATAA

Per esempio, se *E. coli* è tenuto a temperature più alte della temperatura fisiologica di 37 °C, si indurrà la produzione di σ^{32} , che dirige la polimerasi a trascrivere geni che codificano per proteine che proteggono il batterio dallo shock termico. Fattori σ diversi possono essere usati per controllare l'inizio della trascrizione e, in *Bacillus subtilis*, per esempio, dove esistono più di 10 σ differenti, ciò avviene quando questo batterio viene infettato dal batteriofago SPO1. In questo processo infettivo è importante che la trascrizione dei geni del fago avvenga in una sequenza temporale precisa. Prima vengono trascritti i geni precoci, poi i geni intermedi e, da ultimo, i geni tardivi, che sono richiesti per l'assemblaggio finale delle particelle fagiche. Come illustrato in **figura 9.16**, questo controllo temporale è gestito da una serie di fattori σ alternativi. All'inizio l'oloenzima batterico con il fattore σ^{43} di *B. subtilis*, equivalente a σ^{70} di *E. coli*, trascrive sul DNA del fago i geni precoci che avranno le tipiche sequenze -10 e -35. Tra i geni precoci codificati

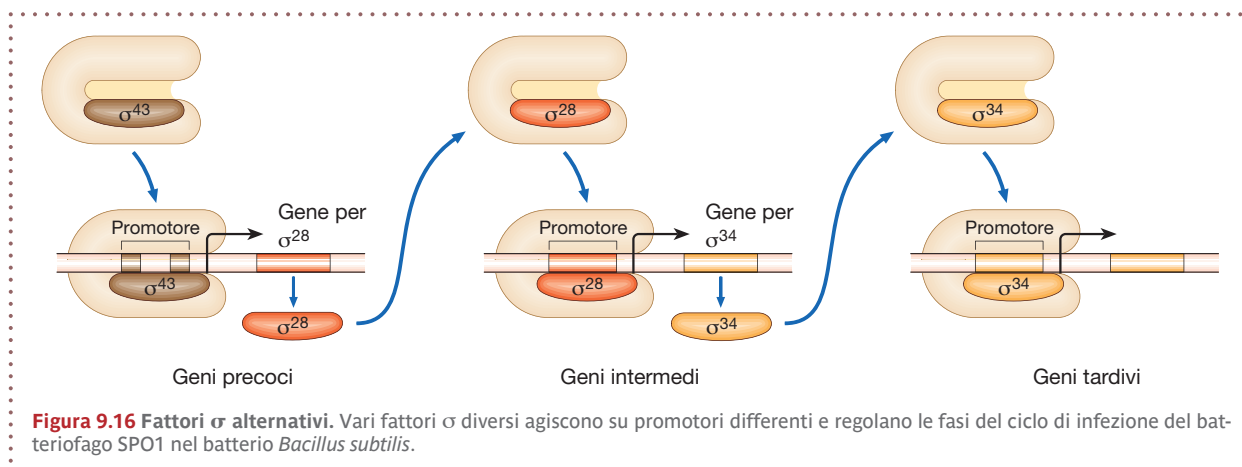


Figura 9.16 Fattori σ alternativi. Vari fattori σ diversi agiscono su promotori differenti e regolano le fasi del ciclo di infezione del batteriofago SPO1 nel batterio *Bacillus subtilis*.

dal fago è presente quello per un nuovo fattore σ che si chiama σ^{28} . Questo si lega alla polimerasi e dirige la trascrizione del gruppo di geni intermedi, tra i quali c'è quello codificante il fattore σ^{34} , specifico per il promotore dei geni tardivi.

9.5 Distacco dell'RNA polimerasi dal promotore e allungamento dell'RNA

Dopo la fase iniziale abortiva, quando l'RNA raggiunge una lunghezza di 9-12 nucleotidi, alcuni cambiamenti di conformazione del fattore σ permettono lo scorrimento dell'RNA polimerasi lungo il DNA e inizia la fase processiva di allungamento del trascritto.

Durante ciascuna fase della trascrizione l'attività dell'RNA polimerasi è modulata dall'interazione con diversi fattori di regolazione e, in particolare, nella fase di allungamento esistono fattori specifici che facilitano questo processo. Nonostante la processività delle RNA polimerasi, l'allungamento del trascritto non è affatto un processo continuo e senza interruzioni. A causa di segnali di regolazione e ostacoli dovuti a particolari sequenze, l'RNA polimerasi spesso rallenta, fa delle pause anche di qualche minuto e a volte si ferma sul DNA. In alcune situazioni le RNA polimerasi possono arretrare, creando dei complessi inattivi dal punto di vista trascrizionale dove il 3'-OH del trascritto si distacca dal sito attivo. Questi complessi che hanno subito un arresto e che sono tornati indietro, possono essere riattivati mediante il taglio idrolitico dell'estremità del trascritto e il ripristino del 3'-OH terminale in registro con il sito catalitico dell'RNA polimerasi. Esiste una classe di fattori di allungamento, come GreB e GreA (e negli eucarioti TFIIS), che aumentano la velocità di allungamento mitigando le pause e riattivando i complessi arrestati. Questo meccanismo, illustrato per GreB in **figura 9.17**, si basa sulla stimolazione dell'attività endogena di editing endonucleolitico della polimerasi, mediante una lunga struttura coiled-coil di GreB che si inserisce nel canale di ingresso dei nucleotidi fino al sito attivo della polimerasi; la funzione di questa viene ripristinata dopo l'eliminazione di un tratto di trascritto che contiene l'errore e che può arrivare fino a 18 nucleotidi.

Quando una RNA polimerasi incontra dei danni sul DNA può anche arrestarsi del tutto; una polimerasi bloccata sul DNA impedisce l'avanzamento di altre polimerasi che stanno trascrivendo lo stesso gene. La rimozione della polimerasi bloccata e il reclutamento degli enzimi per la riparazione, Uvr(A)BC in *E. coli* sono promossi dalla proteina TRCF (*Transcription-Repair Coupling Factor*). Questo enzima è una DNA traslocasi che usa ATP per scorrere lungo il DNA. Una volta raggiunta la polimerasi bloccata, questa viene rimossa dissociando il

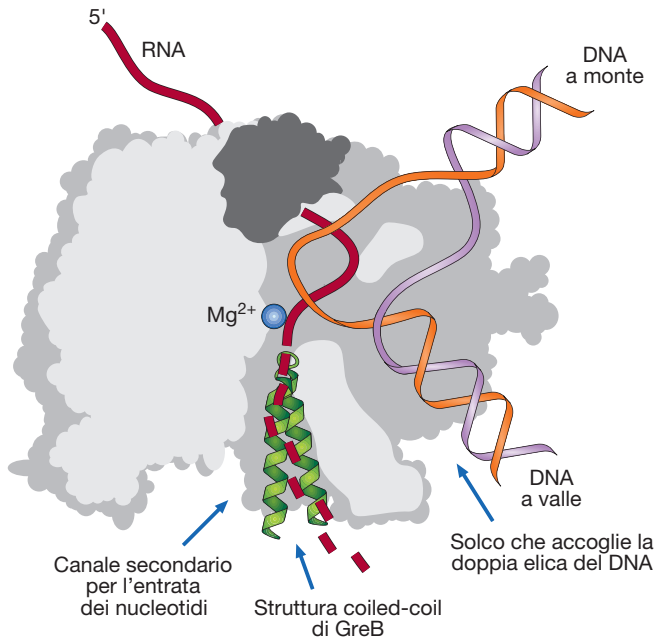


Figura 9.17 Arretramento (back-tracking) dell'RNA polimerasi e controllo dell'allungamento da parte del fattore GreB. La figura mostra uno spaccato dell'RNA polimerasi: sotto è indicato il canale secondario per l'ingresso dei nucleotidi in cui si inserisce il tratto coiled-coil di GreB (verde); sopra a sinistra dell'RNA polimerasi si vede una coda di RNA (rosso) già sintetizzato e il tratteggio rosso che lo prosegue in basso rappresenta il tratto di RNA, liberato dall'arretramento della RNA polimerasi sul DNA, che viene poi rimosso dall'attività endonucleasica della polimerasi stessa. È anche mostrato l'atomo di magnesio vicino al sito attivo (piccola sfera blu) della polimerasi. [Modificata da: Opalka, N. et al. (2003), *Cell*, 8, pp. 335-345.]

complesso di allungamento; TRCF recluta UvrA, che fa parte del macchinario del NER nei batteri, sul DNA danneggiato (vedi finestra 7.1).

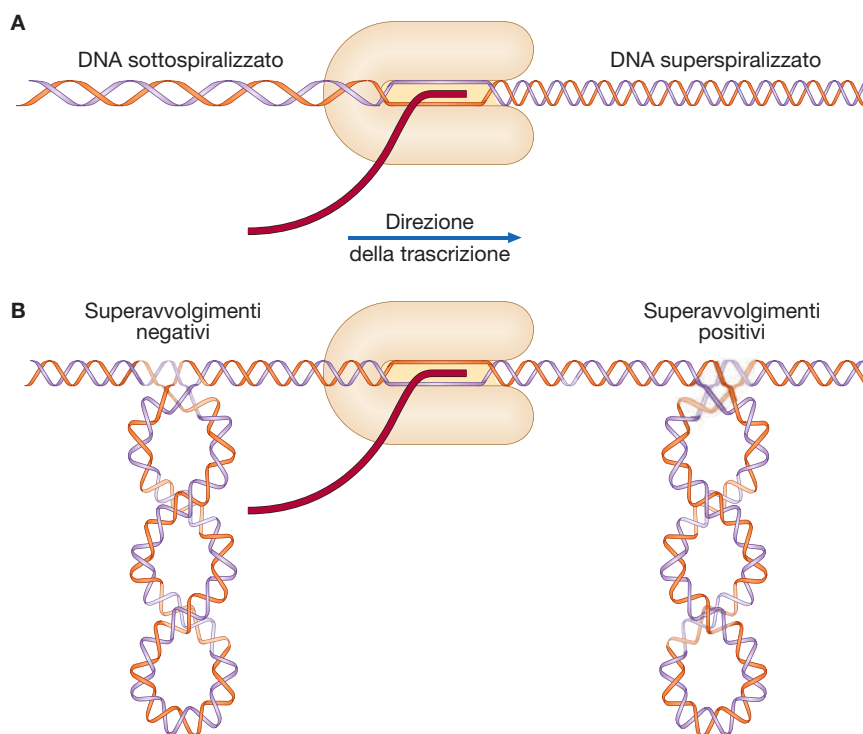
In media durante la trascrizione viene fatto un errore ogni 10^4 nucleotidi, contro un nucleotide ogni 10^7 nella replicazione del DNA.

Topologia e trascrizione

La struttura topologica del DNA ha un ruolo importante nella trascrizione, sia nella fase di inizio che nella fase di allungamento. Come mostrato in **figura 9.18**, durante la fase di allungamento, mentre l'RNA polimerasi avanza lungo la doppia elica, il DNA si apre davanti all'enzima e si richiude dietro. Questo fenomeno genera superavvolgimenti positivi davanti e superavvolgimenti negativi dietro alla polimerasi e viene definito **modello dei domini gemelli per la trascrizione** o **twin domain model**. Ogni 10 nucleotidi che vengono polimerizzati si genera un superavvolgimento positivo a valle e un superavvolgimento negativo a monte. Se si accumulano troppi superavvolgimenti positivi, il DNA diventa troppo compatto e la trascrizione si arresta: infatti, come abbiamo accennato precedentemente, alcune DNA topoisomerasi sono coinvolte nella trascrizione proprio per permettere l'avanzamento dell'RNA polimerasi.

Nel DNA superavvolto negativamente, che è la condizione fisiologica del cromosoma di *E. coli*, risulta facilitata l'apertura dei due filamenti e, quindi, la formazione del complesso aperto. Ci sono dei geni che sono completamente dipendenti dalla presenza di superavvolgimenti negativi perché le loro sequenze consenso sono meno ricche di AT e, quindi, meno facili ad aprirsi per formare la bolla di trascrizione. Possiamo dire che la dipendenza di alcuni promotori dal DNA superavvolto dipende dal contesto generale delle sequenze che fanno parte del promotore. In alcuni casi, un gene il cui promotore può essere attivato solo se il DNA è negativamente superavvolto è posizionato vicino, ma in direzione opposta, a un altro gene in modo che l'attivazione e trascrizione di quest'ultimo, inducendo la formazione di superavvolgimenti negativi a monte, faciliti l'attivazione anche del gene adiacente.

Figura 9.18 Modello dei domini gemelli (*twin domain model*) per la trascrizione. Durante la trascrizione, l'avanzamento della RNA polimerasi lungo il DNA e l'apertura della bolla di trascrizione inducono la formazione di domini superspiralizzati a valle e di domini sottospiralizzati a monte, che vengono chiamati "domini gemelli". Questi possono essere rappresentati come cambiamento del Twist della doppia elica (A) o come formazione di superavvolgimenti dell'asse della doppia elica, Writhe (B). In natura si possono verificare entrambe queste situazioni. Lo stato topologico del DNA è tenuto sotto controllo da alcune topoisomerasi specifiche.



9.6 Terminazione della trascrizione nei procarioti

Abbiamo già detto che la trascrizione termina quando l'RNA polimerasi raggiunge delle sequenze specifiche sul DNA chiamate **terminatori**. L'arresto impedisce l'aggiunta di nuovi nucleotidi, l'enzima si dissocia dal DNA ed è pronto per un nuovo ciclo. È necessario che tutti i legami che mantengono uniti i due filamenti nel tratto ibrido DNA-RNA vengano rotti per permettere il ri-appaiamento tra i filamenti complementari del DNA e la dissociazione del complesso.

In *E. coli* esistono due meccanismi diversi per terminare la trascrizione: in un primo caso, sequenze denominate **terminatori intrinseci** inducono la polimerasi a staccarsi dal complesso e a rilasciare la catena di RNA prodotto, senza l'ausilio di alcun fattore aggiuntivo. Nel secondo caso, denominato **terminazione Rho-dipendente**, la sequenza di terminazione non è in grado di promuovere da sola la dissociazione del complesso, ma necessita dell'azione della proteina Rho.

Terminatori intrinseci

Vediamo ora come sono strutturati i terminatori intrinseci, detti anche **Rho-indipendenti**.

Per indurre la dissociazione del complesso di allungamento, un tipico terminatore intrinseco è rappresentato da un tratto di DNA che contiene delle sequenze palindromiche ricche in G-C seguite da un tratto di 8-9 nucleotidi ricco in A e T. L'RNA trascritto forma, nella regione palindromica ricca in GC, una forcina (hairpin) che destabilizza, insieme con il tratto di 8 U, il complesso trascrizionale. In **figura 9.19** è mostrata la struttura di un terminatore intrinseco con la forcina di RNA, mentre in **figura 9.20** viene presentata una semplificazione degli eventi che portano alla terminazione.

Studi di mutagenesi hanno dimostrato che i legami idrogeno tra le G-C alla base della forcina sono essenziali per la terminazione, come anche il tratto di U che segue. Inoltre, si pensa che la forcina induca la dissociazione della regione ibrida DNA-RNA

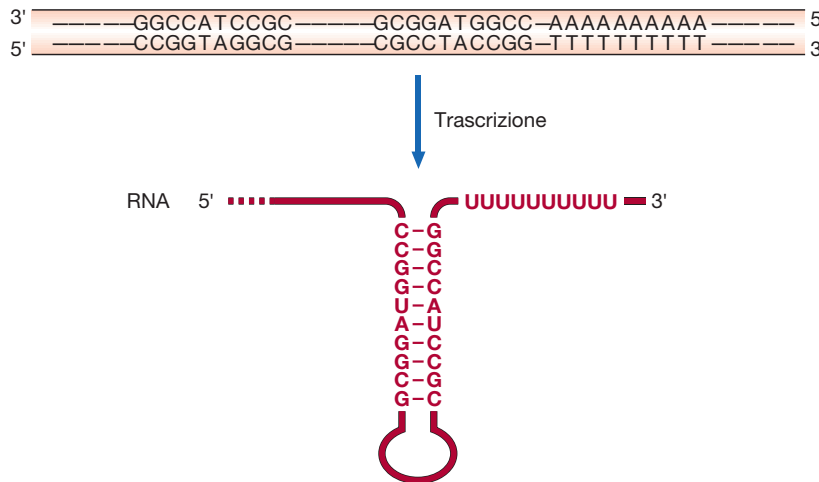


Figura 9.19 Struttura di un terminatore intrinseco. A livello di DNA il terminatore presenta due sequenze ripetute e invertite seguite, sul filamento stampo, da una sequenza di numerose A. L'RNA trascritto assumerà la struttura di un terminatore, cioè una forcina di terminazione seguita da una stringa di U.

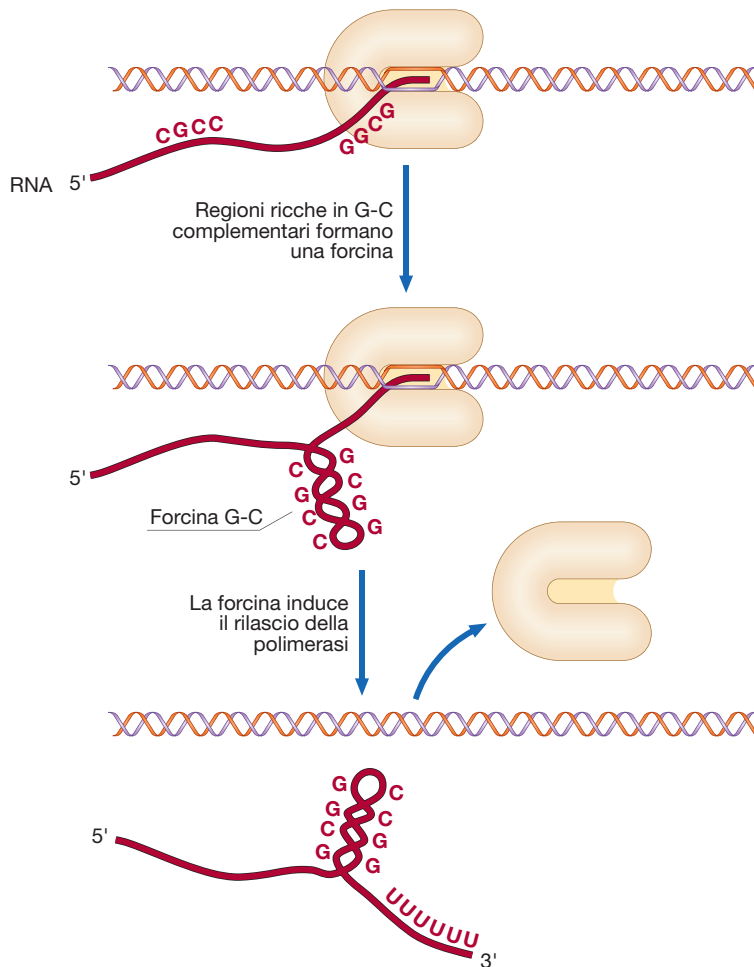


Figura 9.20 Rappresentazione schematica della terminazione intrinseca (Rho-indipendente). La polimerasi incontra e trascrive la regione del terminatore; l'RNA nascente assume la struttura a forcina che destabilizza il complesso di allungamento e ne provoca la dissociazione in corrispondenza della sequenza di U.

e un cambiamento conformazionale di alcuni domini dell'RNA polimerasi che facilitano la dissociazione del complesso. Alcuni studi ipotizzano anche un'influenza dei due atomi di Mg^{2+} , essenziali per la sintesi dell'RNA. La forcina sarebbe simile a un regolatore allosterico che controlla la stabilità del complesso di allungamento. Questo meccanismo di terminazione interessa circa la metà dei geni trascritti in *E. coli*.

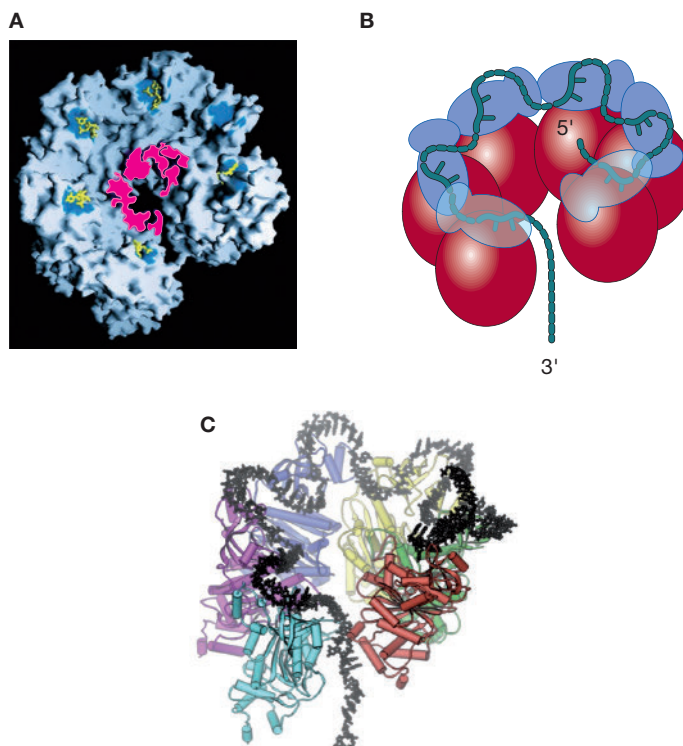
Terminatori Rho-dipendenti

L'altro meccanismo di terminazione è mediato da una proteina che si chiama Rho e che si trova in tutti i batteri. *In vivo* Rho (ρ) si associa all'RNA nascente in una regione ricca di citosine lunga circa 40 nucleotidi, denominata **sito di utilizzo di Rho (rut)**. Questo sito si trova a valle della regione codificante e quindi Rho non incontrerà sul suo cammino un ribosoma che sta sintetizzando la proteina corrispondente. Una volta legato, Rho agisce come una elicasi ATP-dipendente che si muove in direzione 5'→3', per traslocare verso il sito delle trascrizione e promuovere il distacco della polimerasi. Rho può a sua volta essere regolata da altre proteine che interrompono la terminazione in un processo chiamato "anti-terminazione", di cui parleremo in seguito.

La **figura 9.21** mostra una recente struttura tridimensionale di Rho in cui è possibile osservare la sua organizzazione a esamero. In **figura 9.21A** si può vedere la superficie di uno dei lati dell'anello esamerico, dove interagisce l'RNA e, al centro, in rosso, il dominio ATPasico, necessario per la traslocazione. L'RNA si lega a una faccia della molecola e viene inglobato al suo interno durante la traslocazione, come si può vedere in **figura 9.21B**. In **figura 9.21C** si vede la struttura esamerica della proteina. Appena Rho si lega sul trascritto crescente di RNA, il singolo filamento interagisce con il foro dell'esamero e con domini della proteina che lo circondano. Una volta formato il complesso, il legame con l'ATP e la sua idrolisi provocano la traslocazione di Rho lungo l'RNA, fino a che raggiunge la polimerasi e dissocia il complesso e l'ibrido DNA-RNA.

Oltre ai siti "rut", Rho utilizza per terminare la trascrizione delle sequenze di terminazione molto simili ai terminatori intrinseci, ma con una forcina più breve. La **figura 9.22** mostra questa sequenza, che ha una simmetria diadica, e le mutazioni che ne alterano la funzione. La **figura 9.23** illustra un possi-

Figura 9.21 Struttura tridimensionale e organizzazione a esamero di Rho. In (A) si può vedere la superficie di uno dei lati dell'anello esamerico, dove interagisce l'RNA, e al centro, in rosso, il dominio ATPasico, necessario per la traslocazione. L'RNA si lega a una faccia della molecola e viene inglobato al suo interno durante la traslocazione, come si può vedere in (B). In (C) si osserva la struttura tridimensionale della proteina con l'RNA legato alle varie subunità. (Per gentile concessione di J. Berger, Johns Hopkins University.)



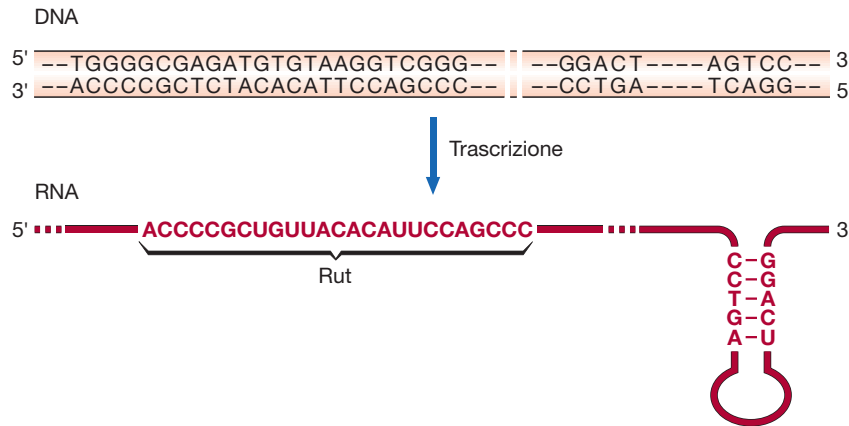


Figura 9.22 Struttura dei terminatori Rho-dipendenti. A livello del DNA il terminatore Rho-dipendente è costituito da due brevi sequenze ripetute e invertite (a simmetria diadica o palindromica) precedute, a una distanza variabile, dalla sequenza Rut, ricca in C sul filamento codificante. Conseguentemente, l'RNA nascente presenterà la sequenza rut seguita, a una certa distanza, da una breve forcina.

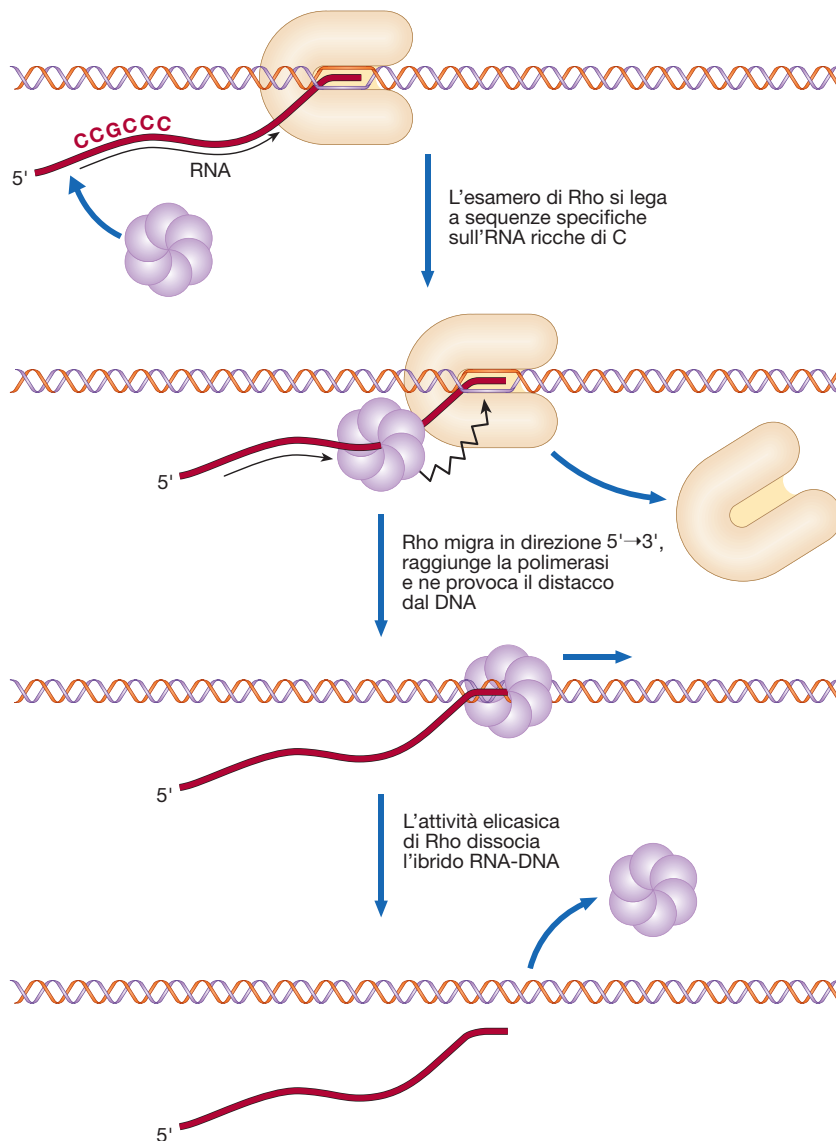


Figura 9.23 Modello per la terminazione Rho-dipendente. Quando l'RNA polimerasi ha trascritto e superato rut, questo lega il fattore Rho, che poi trasloca lungo l'RNA seguendo la polimerasi. Quando questa sintetizza la sequenza diadica, che assume una struttura a forcina, rallenta (o stalli), Rho la raggiunge e innesca la terminazione.

bile modello della terminazione Rho-dipendente dove si vede l'enzima che si lega al sito "rut", trasloca lungo l'RNA, raggiunge la polimerasi, la dissocia e, infine, la sua attività elicastica dissocia l'ibrido RNA-DNA, rilasciando il messaggero.

La terminazione della trascrizione è una fase in cui si realizzano alcuni importanti meccanismi di regolazione dell'espressione genica. Infatti la modulazione dell'efficienza del terminatore di un gene ha effetti sull'espressione dei geni che si trovano adiacenti a valle. Presenteremo due esempi di questo tipo di regolazione, l'antiterminazione e l'attenuazione, nel prossimo capitolo (*vedi* par. 10.3) sulla regolazione della trascrizione.

Lecture di approfondimento

Campbell, E.A., Westblade, L.F. e Darst, S.A. (2008), «Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective», *Current Opinion in Microbiology*, **11**(2), pp. 121-127.

Murakami e Darst (2003), «Bacterial RNA polymerases: the whole story», *Current Opinion in Structural Biology*, **13**(1), pp. 31-39.

Roberts, J.W. (2006), «RNA polymerase, a scrunching machine», *Science*, **314**(5802), pp. 1097-1098.

Skordalakes, E. e Berger, J.M. (2003), «Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading», *Cell*, **114**(1), pp. 135-146.

10

Regolazione della trascrizione nei procarioti

10.1 Elementi di controllo nella trascrizione dei procarioti

10.2 Operone *lac*

10.3 Operone del triptofano

10.4 Controllo dell'espressione genica nel fago lambda

Nella gran parte dei procarioti l'espressione genica è regolata in modo da assicurare alla cellula un adattamento a cambiamenti fisici o nutrizionali dell'ambiente in cui vive; infatti, molti meccanismi regolativi determinano variazioni dei livelli di espressione dei geni in relazione alle richieste della cellula connesse alla sua fase di crescita e di sviluppo. Si è visto, per esempio, che una variazione nella temperatura di crescita modifica l'espressione di 50-100 geni.

La regolazione dell'espressione genica può essere attuata a livello di ciascuno dei processi che, a partire dal DNA, portano al prodotto genico finale: dalla trascrizione, al processamento del trascritto, al processo di traduzione e all'assemblaggio di polipeptidi nella molecola proteica funzionale. Benché la via verso i prodotti funzionalmente attivi possa essere regolata a diversi livelli, la regolazione nei procarioti avviene principalmente a livello trascrizionale, modulando, in particolare, la fase di inizio della trascrizione.

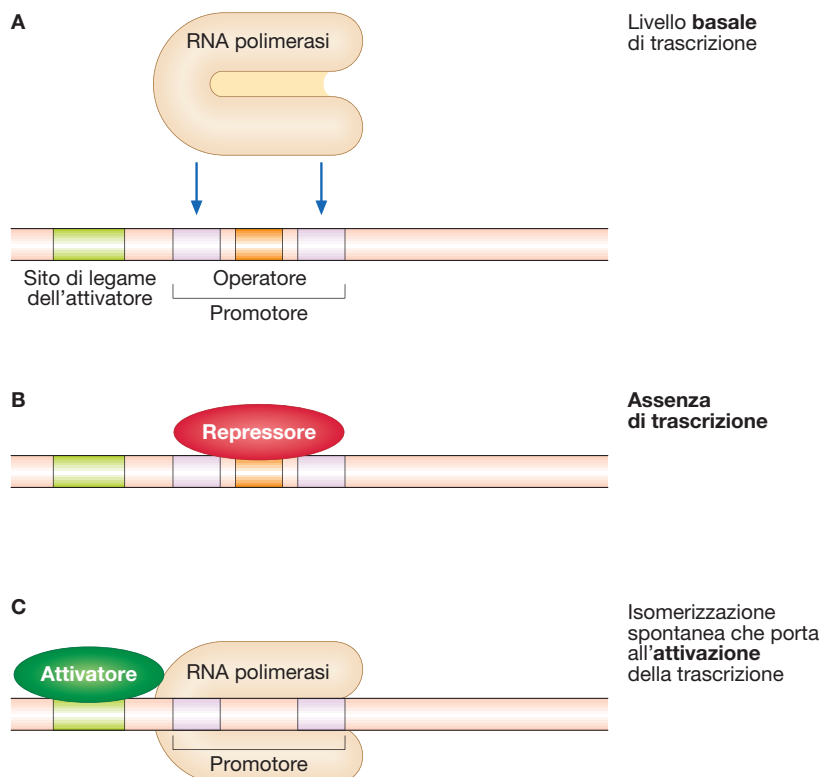
In questo capitolo saranno principalmente descritti i meccanismi che nei batteri regolano in modo specifico l'inizio della trascrizione.

10.1 Elementi di controllo nella trascrizione dei procarioti

Attivatori trascrizionali, repressori e operatori

Le cellule procariotiche sono particolarmente sensibili alla presenza di segnali esterni, per lo più molecole presenti nel terreno di coltura che entrano nella cellula e interagiscono con proteine regolatrici sia di tipo positivo che negativo. Le prime, chiamate **attivatori**, sono implicate nella **regolazione positiva** contribuendo a incrementare il livello di trascrizione, mentre le seconde, dette **repressori**, sono implicate nella **regolazione negativa** inibendo la trascrizione. Queste proteine regolatrici agiscono a livello della prima fase della trascrizione. Infatti, legandosi a specifici siti in prossimità della sequenza codificante di un gene, influenzano il legame dell'RNA polimerasi al promotore: gli attivatori favoriscono tale legame, mentre i repressori lo inibiscono. In genere, in assenza di proteine regolatrici, il legame della RNA polimerasi al promotore è debole e risulta in un livello di espressione costitutiva **basale**. Questa trascrizione risulta inibita in

Figura 10.1 La regolazione della trascrizione può essere positiva e negativa. (A) L'RNA polimerasi è in grado di legarsi alle sequenze del promotore trascrivendo a bassi livelli l'RNA. (B) *Controllo negativo*: in presenza di una molecola di repressore che si lega al sito operatore, sovrapposto alle sequenze del promotore, la polimerasi non è più in grado di legarsi e non c'è trascrizione. (C) *Controllo positivo*: la polimerasi viene reclutata sul promotore con grande efficienza da una molecola di attivatore che si lega nella regione del promotore a monte dei siti di riconoscimento per la polimerasi. In questo caso i livelli di RNA prodotto sono molto alti.



seguito al legame del repressore su una sequenza del DNA che prende il nome di **operatore**, sovrapposta o adiacente al promotore cui si lega l'RNA polimerasi. L'attivatore, al contrario, ha due domini di legame, uno interagisce con l'RNA polimerasi, l'altro riconosce una sequenza sul DNA vicino al promotore, promuovendo un legame stretto e cooperativo della proteina sul DNA e scatenando la spontanea isomerizzazione in una conformazione aperta che favorisce l'inizio della trascrizione (**fig. 10.1**).



Figura 10.2 François Jacob e Jacques Monod. Jacob, a sinistra, e Monod, a destra, in una foto del 1971 nel loro laboratorio all'Istituto Pasteur a Parigi, hanno ricevuto, nel 1965, il premio Nobel per la medicina per il loro lavoro sull'organizzazione e regolazione degli operoni batterici.

Organizzazione degli operoni procariotici

Nel genoma dei procarioti i geni sono per lo più organizzati in operoni. Un **operone** è un gruppo di geni adiacenti trascritti insieme in una singola molecola di mRNA che per questo viene chiamato **policistronico**. Gli operoni, con un numero di geni variabile dalla singola unità fino a più di una dozzina, sono specifici degli organismi procariotici. Come vedremo nel capitolo 11, negli eucarioti i geni non sono organizzati in operoni, ma sono trascritti singolarmente. Tuttavia, a questa regola ci sono alcune eccezioni; per esempio, in alcuni particolari gruppi di eucarioti (ad es. nematodi e ascidie) alcuni geni possono essere organizzati come negli operoni procariotici e sono espressi come trascritti policistronici e poi risolti in mRNA monocistronici attraverso il meccanismo del trans-splicing (*vedi* par. 13.1). Altre eccezioni le troviamo in alcuni mRNA policistronici virali.

Nella gran parte dei casi i geni di un operone, oltre a essere disposti sequenzialmente nel genoma, sono anche funzionalmente correlati in quanto codificano proteine implicate in una comune via (*pathway*) metabolica. Di conseguenza, la regolazione della trascrizione, agendo sull'intero gruppo genico, ne coordina l'espressione. La struttura dell'operone fu descritta per la prima volta nel 1961 dai due biologi francesi François Jacob e Jacques Monod (**fig. 10.2**), sulla base di studi

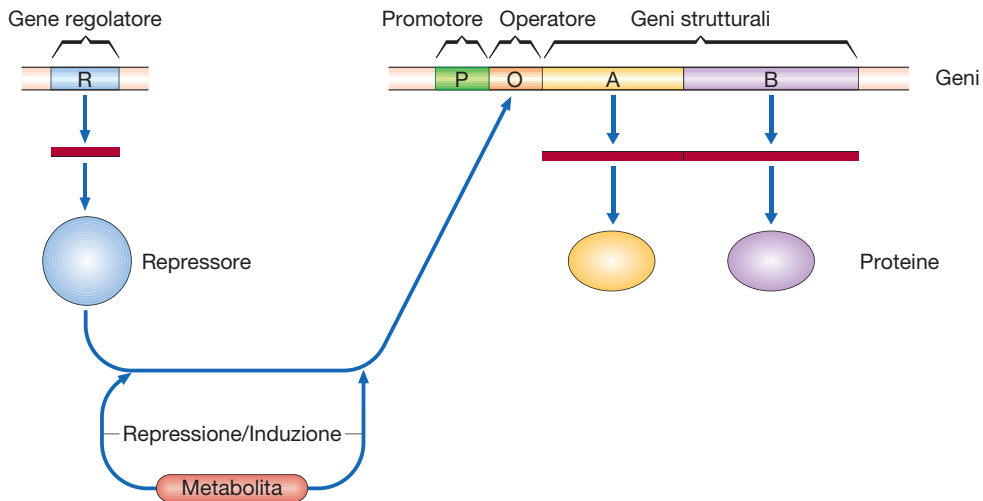


Figura 10.3 Nei procarioti i geni sono organizzati in operoni. Un tipico operone è costituito da: uno o più geni strutturali adiacenti che codificano proteine (A e B nella figura); una sequenza definita promotore (P) che lega l'RNA polimerasi a monte della serie di geni

strutturali; una sequenza definita operatore (O) che lega una proteina regolatrice che interferisce con il legame dell'RNA polimerasi; un gene regolatore (R) che codifica per la proteina regolatrice.

genetici e biochimici riguardanti l'effetto di mutazioni che influenzano l'utilizzo del lattosio come fonte energetica in *Escherichia coli* (finestra 10.1). Gli operoni contengono, oltre ai geni strutturali che devono essere trascritti, delle sequenze di controllo che ne regolano l'espressione. Come mostrato in figura 10.3, un tipico operone è costituito da: (1) uno o più **geni strutturali** adiacenti che codificano proteine; (2) un'unica sequenza **promotore** a monte della serie di geni capace di legare l'RNA polimerasi; (3) una sequenza **operatore** sovrapposta o adiacente al promotore che, interferendo con il legame della polimerasi al promotore, regola l'espressione dei geni strutturali; infine (4) un **gene regolatore** che codifica per la proteina regolatrice ma non è considerato parte integrante dell'operone, in quanto può essere dislocato in un punto del genoma lontano da esso.

10.2 Operone *lac*

A partire dal 1961, l'**operone *lac*** è considerato il modello paradigmatico di regolazione genica nei procarioti. Esso è soggetto a due meccanismi di regolazione: (1) una **regolazione specifica**, realizzata mediante un controllo negativo da parte di un repressore e che risponde alla disponibilità di lattosio; (2) una **regolazione globale**, di cui parleremo più avanti, realizzata da un controllo positivo da parte di una proteina attivatrice e che risponde a cambiamenti ambientali più ampi come il fabbisogno di energia da parte della cellula.

Controllo negativo dell'operone *lac*

Per quanto riguarda la prima regolazione, essa è responsabile del controllo della sintesi degli enzimi necessari quando la cellula si trova a dovere utilizzare il lattosio (fig. 10.4) come sorgente di energia; quando questi enzimi non sono necessari la loro sintesi viene "spenta". Il metabolismo del lattosio (fig. 10.4) necessita di due enzimi, la **β -galattosidasi** e la **permeasi**, codificati dai due geni strutturali contigui *lacZ* e *lacY* (fig. 10.5). *lacZ* codifica per l'enzima β -galattosidasi che scinde il disaccaride lattosio in glucosio e galattosio (fig. 10.4). Nella sua conformazione enzimatica attiva è un tetramero di 500 kDa. *lacY* codifica per una permeasi che

FINESTRA 10.1 **GLI ESPERIMENTI DI JACOB E MONOD**

Tutto quello che oggi sappiamo sulla regolazione dell'attività genica nei procarioti, ma anche negli eucarioti, ha come fondamento alcuni concetti generali definiti dai due ricercatori francesi François Jacob e Jacques Monod con una serie di esperimenti molto eleganti pubblicati in un famoso lavoro del 1961. Jacob e Monod lavoravano all'Istituto Pasteur di Parigi, dove vi era molto interesse per le fermentazioni, cioè l'utilizzo di colture di microrganismi per la produzione di molecole quali antibiotici, proteine e altro. Si era capito che la produzione di queste molecole è molto influenzata dalla composizione del mezzo di coltura, in quanto l'attività dei geni può essere regolata da molecole immesse nel mezzo.

La crescita di batteri quale *Escherichia coli* dipende dalla presenza nel mezzo di coltura di almeno uno zucchero che funge da fonte di energia e di carbonio. Gli zuccheri che possono essere utilizzati sono vari: glucosio, lattosio, arabinosio, maltosio, raffinosio e altri. Benché la cellula batterica sia in grado di produrre gli enzimi per utilizzare ciascuno di questi zuccheri, in realtà in ogni momento produce solo quelli relativi allo zucchero, o agli zuccheri, effettivamente presente/i nel mezzo, evitando così di sprecare energia per sintetizzare gli enzimi per altri zuccheri. Questo tipo di regolazione è stato appunto affrontato da Jacob e Monod, che scelsero come modello sperimentale il batterio *E. coli* e, in particolare, la regolazione della sintesi della β -galattosidasi, l'enzima necessario per utilizzare il lattosio. Poiché la sintesi di questo enzima viene indotta nella cellula dall'aggiunta al mezzo di coltura di lattosio, questo viene chiamato **induttore**. Se batteri provenienti da una coltura contenente glucosio vengono trasferiti in un mezzo fresco pure contenente glucosio, essi continuano a crescere. Se invece vengono trasferiti in un mezzo contenente non più glucosio ma lattosio, essi smettono di crescere per un breve periodo per poi riprendere dopo pochi (5-10) minuti. Questo ritardo è dovuto al fatto che per potere utilizzare il lattosio è necessario che il batterio sintetizzi gli enzimi necessari che non sono presenti nella cellula cresciuta in assenza di lattosio. In altre parole, al momento del cambiamento di zucchero il batterio attiva i geni per sintetizzare alcuni nuovi enzimi e disattiva i geni per gli enzimi che non sono più necessari. Questa attivazione richiede alcuni minuti di tempo affinché i nuovi geni siano attivati e trascritti, e

gli mRNA tradotti in proteine. Volendo studiare questo meccanismo di regolazione Jacob e Monod, che erano essenzialmente dei genetisti, decisero di isolare e studiare dei mutanti. Attenzione! Non mutanti dei geni strutturali per la β -galattosidasi, ma mutanti della sua regolazione, cioè batteri che producono questo enzima sempre, anche quando cresciuti in assenza di lattosio. Il fenotipo di questi batteri consiste nel fatto che, in un esperimento di cambiamento di zucchero come quello sopra descritto, dopo il trasferimento nel mezzo con lattosio i batteri non presentano il caratteristico ritardo della crescita. Infatti, essi hanno già l'enzima β -galattosidasi in quanto la sua sintesi non è regolata. In questa maniera Jacob e Monod isolarono numerosi di questi **mutanti costitutivi** (*lac^C*), così chiamati perché producono β -galattosidasi sempre, in modo cioè costitutivo. Quando andarono a fare un'analisi genetica di questi mutanti si accorsero che, pur presentando tutti lo stesso fenotipo, in realtà potevano essere distinti in due categorie diverse. Una prima categoria comprendeva mutanti che se messi in parziale diploidia sono dominanti rispetto al wt; inoltre, queste mutazioni mappano in posizione adiacente al gene strutturale *lacZ* che codifica per la β -galattosidasi. Le mutazioni della seconda categoria al contrario, se messe in parziale diploidia, sono recessive rispetto al gene normale (wt); queste mutazioni mappano più a monte, a una certa distanza dal gene *lac*. Questi dati sono essenzialmente quelli che hanno permesso a Jacob e Monod di proporre il **modello dell'operone**. Infatti, l'essenza di questo modello è che la regolazione del gene *lac* è dovuta a due loci (sequenze). Uno è adiacente ai geni strutturali *lac* e influenza solo questi (*cis*-dominante), l'altro è il gene regolatore che produce il repressore, una molecola diffusibile che può agire anche a distanza, anche su un'altra molecola di DNA (*trans*-dominante). Questo modello, che è stato poi confermato a livello molecolare (vedi testo), non solo è valido per la regolazione della trascrizione nei procarioti, ma possiamo considerarlo nella sua essenza alla base di tutta la regolazione dell'attività genica in procarioti ed eucarioti. Infatti essenzialmente in tutti i meccanismi di regolazione, anche a livello post-trascrizionale (maturazione e traduzione), il controllo è sempre esercitato da geni regolatori che producono molecole *trans*-agenti che regolano i geni bersaglio legandosi a elementi *cis*-agenti.

ha un peso molecolare di circa 30 kDa: questa, legandosi alla membrana, permette l'ingresso all'interno della cellula del lattosio, e in genere dei β -galattosidi. Inoltre, adiacente a questi due, è presente un terzo gene, *lacA*, che codifica per la tiogalattoside transacetilasi, che trasferisce un gruppo acetilico ai β -galattosidi.

Come già osservato (**fig. 10.5**), i tre geni strutturali per gli enzimi LacZ, LacY e LacA sono tra loro adiacenti e costituiscono l'operone *lac*. L'operone si estende per circa 6000 pb ed è caratterizzato da un suo promotore e un suo terminatore della trascrizione; viene così trascritto come un unico mRNA policistronico che porta alla sintesi coordinata di tutti e tre gli enzimi codificati dall'operone. A monte dei geni strutturali vi sono le sequenze regolatrici **P** (promotore) e **O** (operatore) e, ancora più a monte, il gene regolatore *lacI* costitutivamente trascritto a partire da un suo specifico promotore. Quest'ultimo, che può funzionare anche se spostato in una regione lontana del genoma, codifica per il repressore

che, se legato all'operatore O, inibisce la trascrizione dei geni strutturali. Infatti, il repressore presenta due differenti domini di legame: uno riconosce e lega le sequenze dell'operatore, situate 17-25 nucleotidi a monte del gene strutturale *lacZ*, l'altro lega molecole segnale, quali il lattosio e suoi analoghi, che inducono la trascrizione. Il promotore P subito a monte dell'operatore O è costituito come tutti i tipici promotori procariotici dalle sequenze -10 e -35, rappresenta il sito di attacco della RNA polimerasi e regola l'inizio della trascrizione.

Come funziona l'operone *lac*? Nel caso particolare dell'operone *lac* la piccola molecola capace di indurre la trascrizione non è il lattosio, ma il suo analogo **allolattosio** (fig. 10.4), sottoprodotto della β -galattosidasi, che va considerato quindi il vero naturale "induttore" dell'operone anche se, per semplicità, in alcuni casi si parla del lattosio come induttore. In assenza di allolattosio la trascrizione dei geni dell'operone *lac* è inattivata dal repressore che, legandosi all'operatore, impedisce la trascrizione dei geni strutturali. L'aggiunta dell'induttore risulta in un suo legame al repressore che subisce una transizione allosterica. Questo cambiamento conformazionale determina una ridotta affinità del repressore per la sequenza dell'operatore, dal quale perciò si distacca. L'RNA polimerasi è così libera di iniziare la trascrizione dei geni strutturali. Quando l'induttore allolattosio è presente nel mezzo, produce un aumento di β -galattosidasi fino a oltre un migliaio di volte superiore rispetto a condizioni di repressione. La stessa funzione di induttore può essere svolta da altri analoghi del lattosio, tra cui il composto sintetico **IPTG (isopropiltiogalattoside, fig. 10.4)**. Questo è definito come un induttore gratuito in quanto è in grado di legare il repressore e indurre la trascrizione dei geni *lacZ*, *Y* e *A*, ma non è riconosciuto e scisso dalla β -galattosidasi.

Vediamo ora qual è l'effetto dei diversi tipi di mutazioni che sono state trovate per la regolazione della sintesi della β -galattosidasi. Non consideriamo qui i mutanti incapaci di utilizzare il lattosio perché non esprimono una corretta β -galattosidasi essendo mutati, per esempio, nel gene strutturale o nel promotore.

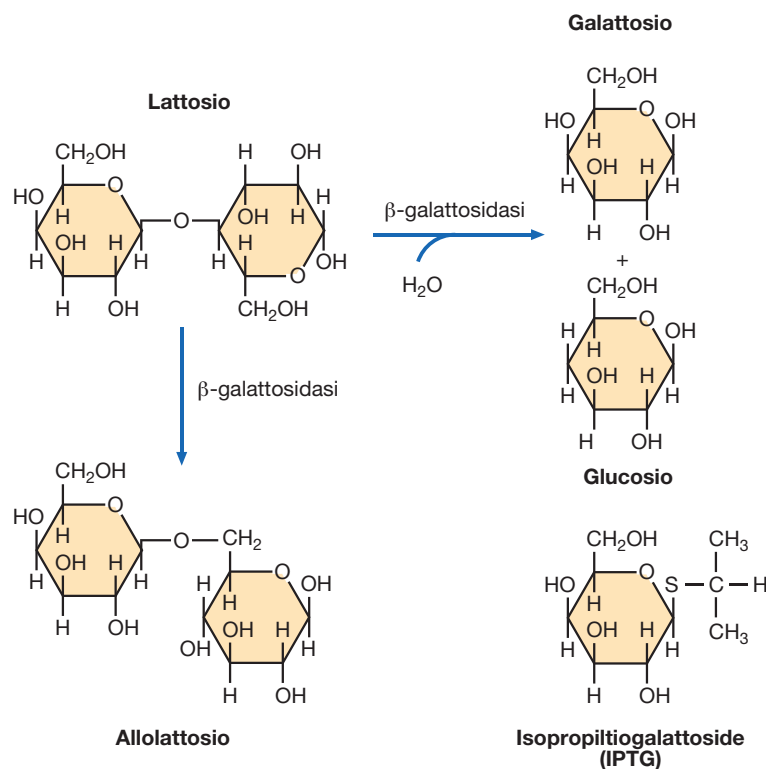


Figura 10.4 Lattosio e β -galattosidasi. Il lattosio può essere modificato dalla β -galattosidasi in allolattosio, che ha una struttura diversa che si lega al repressore, oppure scisso in glucosio e galattosio. L'isopropiltiogalattoside (IPTG) è una molecola che mima l'allolattosio e funziona da induttore per attivare la trascrizione.

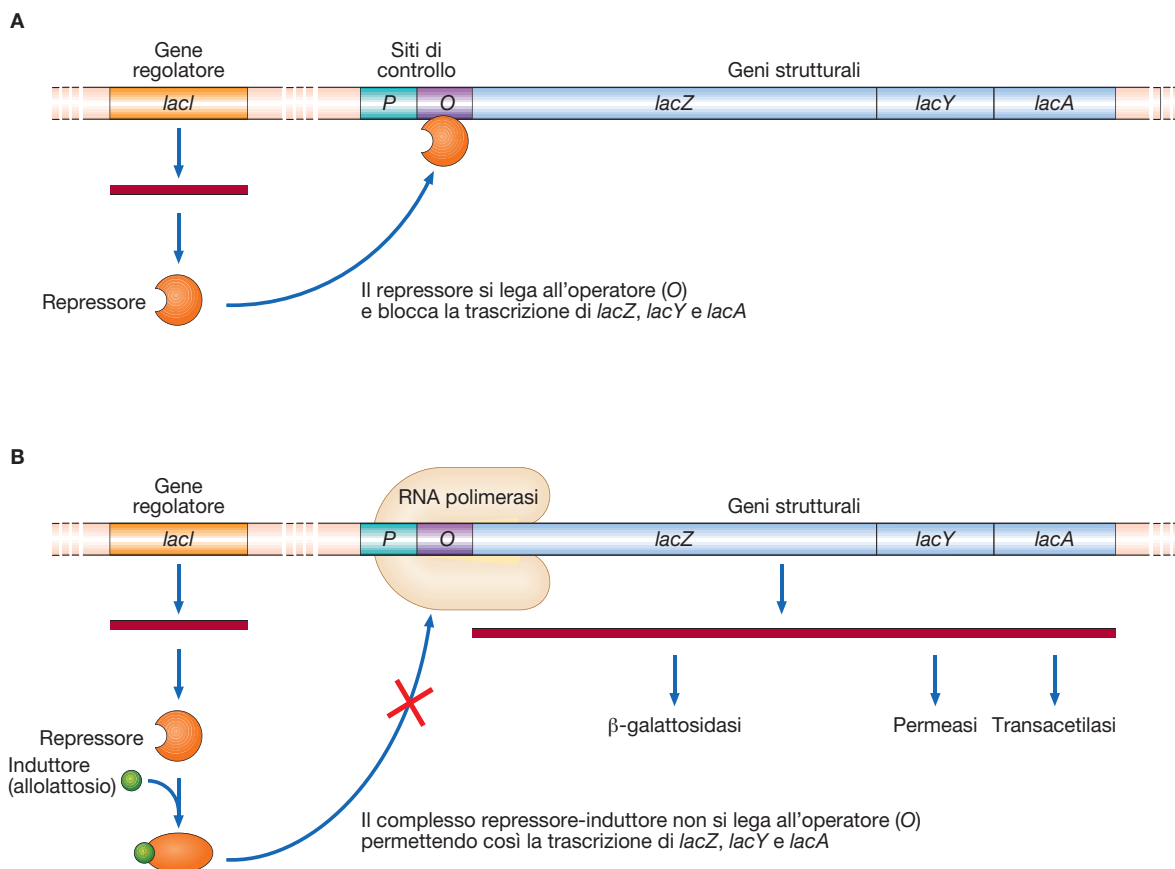


Figura 10.5 Operone lattosio. L'operone lattosio è un operone inducibile a controllo negativo. In (A) è mostrato lo stato represso dell'operone quando, in assenza di lattosio, una molecola di repressore è legata al sito operatore e impedisce alla RNA polimerasi di legarsi al promotore e iniziare la trascrizione dei geni a valle.

In (B) si vede che l'induttore, l'allolattosio o l'IPTG, si lega al repressore, formando un complesso che non è più in grado di legarsi all'operatore. In questo caso, la polimerasi troverà il promotore libero e potrà trascrivere i geni strutturali.

Prendiamo in considerazione invece i **mutanti costitutivi** (Lac^c) che producono una β-galattosidasi funzionale, ma non più regolata dalla presenza/assenza dell'induttore. Di questi ce ne sono due categorie (fig. 10.6), identificati e studiati da Jacob e Monod (finestra 10.1), che mappano rispettivamente nell'operatore O (**mutazioni O^c**) e nel gene regolatore *I* (**mutazioni I^c**). I mutanti delle due categorie presentano lo stesso fenotipo in quanto, in entrambi i casi, è inibito il legame del repressore all'operatore. Invece, se messi in parziale diploidia con una copia di tipo selvatico localizzata su un plasmide, essi presentano fenotipi diversi, rispettivamente costitutivo e wt. Infatti, le mutazioni O^c interessano la sequenza nucleotidica dell'operatore che si trova adiacente ai geni strutturali (fig. 10.6A). Esse hanno permesso di individuare l'operatore come un elemento che non funziona attraverso un prodotto genico, ma solo in quanto capace di essere riconosciuto e legato dal repressore. Se la mutazione impedisce questa interazione, i geni strutturali adiacenti saranno espressi costitutivamente, mentre quelli localizzati sul plasmide restano regolati; comunque nel complesso il batterio presenterà sintesi costitutiva di β-galattosidasi. Poiché l'operatore controlla solo i geni strutturali che si trovano adiacenti e a valle, viene definito come **dominante in *cis*** (o *cis*-dominante). Diversa è la situazione per le mutazioni della seconda categoria, $lacI^c$ (fig. 10.6B). Queste riguardano il gene regolatore *lacI* e risultano nella mancata sintesi di un repressore attivo capace di riconoscere e legare l'operatore. Tuttavia,

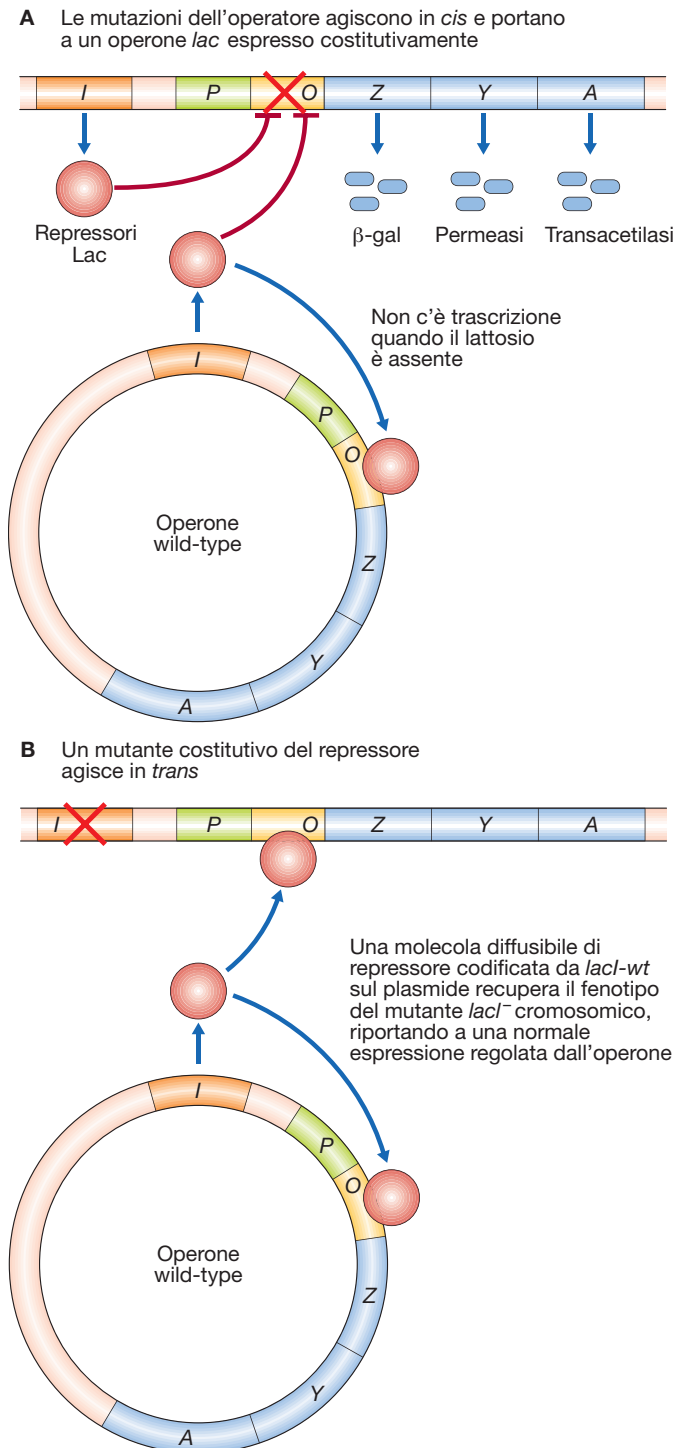


Figura 10.6 L'operatore funziona in *cis*, mentre il repressore agisce in *trans*. Utilizzando cellule di *Escherichia coli* parzialmente diploidi per l'operone *lac*, di cui una copia è localizzata nel cromosoma batterico e l'altra in un plasmide, è stato possibile, mediante analisi di mutazioni specifiche, comprendere il funzionamento dei siti operatori e del repressore. (A) Un batterio con una mutazione *O*^c dell'operatore di una delle due copie dell'operone, presenta un'espressione costitutiva della β -galattosidasi; infatti il repressore, prodotto da ambedue le copie del gene *lacI*, è in grado di legare l'operatore wt ma non l'operatore mutato; si dice che questa mutazione è *cis*-dominante. (B) Un batterio con una mutazione *I*^c di una delle due copie del gene regolatore *lacI*, quella cromosomica nell'esempio mostrato, presenta un'espressione regolata della β -galattosidasi; infatti il repressore prodotto dalla copia plasmidica wt di *lacI* agisce su ambedue gli operatori. Questo dimostra che il repressore agisce in *trans*.

poiché la copia wt del gene *lacI* presente sul plasmide è in grado di legarsi anche all'operatore dell'operone cromosomico, nel complesso il batterio presenterà sintesi regolata di β -galattosidasi. Poiché il gene regolatore controlla, tramite il repressore, i geni strutturali anche se distanti sul cromosoma o anche su un altro cromosoma, viene definito come **dominante in *trans*** (o *trans*-dominante).

Vedremo più avanti (vedi fig. 10.8) altri tipi di mutazioni del gene regolatore, tra cui le mutazioni *lacI*^S, non inducibili; i relativi ceppi mutanti non esprimono

la β -galattosidasi neanche in presenza di induttore. In questo caso la mutazione modifica il dominio di legame dell'induttore. Il repressore così mutato, non essendo più riconosciuto e legato dall'induttore, resta sempre legato all'operatore inibendo permanentemente la trascrizione, anche in presenza di allolattosio.

Va detto che in realtà negli esperimenti di parziale diploidia, in cui un plasmide inserito nel batterio contiene un'altra copia dell'operone del lattosio, la situazione è ulteriormente complicata dal fatto che il repressore è una proteina che in stato attivo si presenta sotto forma di tetramero. Nel caso in cui i due alleli del gene *lacI* siano differenti, per esempio uno mutato e uno wt, i diversi monomeri vanno ad assemblarsi in tetrameri composti da monomeri diversi in varie combinazioni. Questo può portare a un effetto di dominanza negativa in quanto il tetramero può essere incapace di legare l'operatore sebbene la subunità difettosa sia solo una delle quattro (*vedi* par. 20.11).

Repressore Lac e sue interazioni con il promotore

Come illustrato in **figura 10.7**, il repressore Lac lega il sito promotore in forma di dimerico. Questo è costituito da due lunghi monomeri, ciascuno di 38 kDa, strettamente connessi tra loro. In ciascun monomero si riconoscono:

- un dominio N-terminale o “testa” che lega il DNA;
- un dominio centrale “core” che contiene il sito di legame per l'induttore;
- una regione C-terminale di oligomerizzazione, contenente una α -elica implicata nella formazione del tetramero (*vedi* oltre).

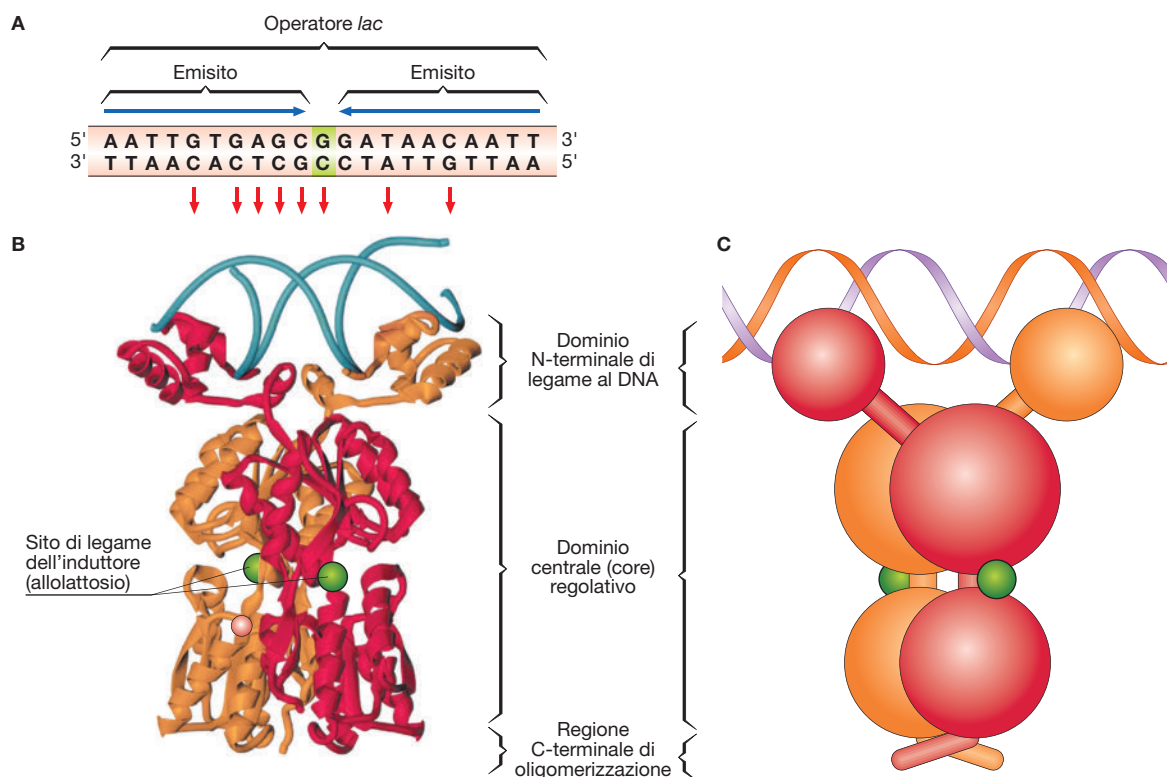


Figura 10.7 Interazione tra repressore e operatore. (A) Ciascun sito di riconoscimento dell'operatore è formato da due “emisiti”, regioni uguali e invertite con sequenza palindromica, e viene legato da un dimerico di repressore. (B) Rappresentazione strutturale del repressore dimerico legato a un sito del repressore in cui ciascun monomero interagisce con un emisito nel solco maggio-

re della doppia elica. (C) Rappresentazione schematica di quanto mostrato in (B). In (B) e (C) sono indicati: il dominio N-terminale di interazione con il DNA, la regione C-terminale implicata nella dimerizzazione e il dominio centrale (core) regolativo che contiene il sito riconosciuto e legato dall'induttore (allolattosio), mostrato come una sferetta verde.

Il motivo strutturale alla base del legame del repressore con il DNA è una struttura secondaria conservata nota come **elica-giro-elica** (*helix-turn-helix*). Come vedremo più a fondo nel prossimo capitolo (*vedi par. 11.7 e fig. 11.20*), il dominio elica-giro-elica è costituito da due α -eliche una delle quali, chiamata R (elica di riconoscimento), riconosce specificamente e lega una regione del DNA nel solco maggiore. La seconda α -elica stabilizza l'interazione tra proteina e DNA, ma non ha un ruolo nella specificità del riconoscimento.

Un dimerico di repressore Lac lega il sito dell'operatore, costituito da 21 pb, la cui sequenza è composta da due parti ripetute (anche se non perfettamente) e invertite come mostrato in **figura 10.7A**. Ciascuno dei due emisiti è riconosciuto da un monomero del repressore (**fig. 10.7B e C**). È chiaro come il riconoscimento e l'interazione tra ciascun monomero e ciascuno dei due emisiti siano simili. Sono tuttavia note alcune distorsioni nel complesso proteina-DNA, come una torsione del DNA nella regione dell'operatore, o interazioni con regioni situate al di fuori dell'operatore che saranno approfondite in seguito.

In **figura 10.8** vengono associate, a ciascun dominio strutturale, le diverse mutazioni del repressore che abbiamo visto nel paragrafo precedente. Le **mutazioni $lacI^S$** impediscono l'interazione del repressore con l'induttore; di conseguenza, anche l'induzione dell'operone è inibita. Mutazioni di questo tipo sono state mappate sia nella tasca dove si lega l'induttore, sia sulla superficie all'interfaccia tra i due dimeri. Queste ultime bloccano la trasmissione delle perturbazioni di struttura in seguito al legame dell'induttore e impediscono il legame con il DNA. Le **mutazioni $lacI^-$** influenzano l'interazione tra i due dimeri e il raggiungimento della conformazione tetramerica e mappano su siti nel dominio centrale o sulla α -elica di oligomerizzazione impedendo, rispettivamente, la formazione del dimerico e l'organizzazione in tetrameri. Infine, il gruppo di **mutazioni $lacI^{-d}$** si trova sul sito di legame per il DNA e determina una mancata attività repressiva.

Come già detto, la sequenza di DNA alla quale si lega il repressore è (quasi) palindromica e ciascuna delle due ripetizioni invertite è riconosciuta dal dominio N-terminale, o "testa", di un monomero di repressore. È evidente come vi sia una corrispondenza tra la sequenza simmetrica dell'operatore e la struttura dimerica del repressore (**fig. 10.7**), corrispondenza che è molto rilevante nell'interazione tra DNA e repressore. Infatti, la testa del repressore monomericico è in grado di legare in modo indipendente mezzo sito del promotore, risultando

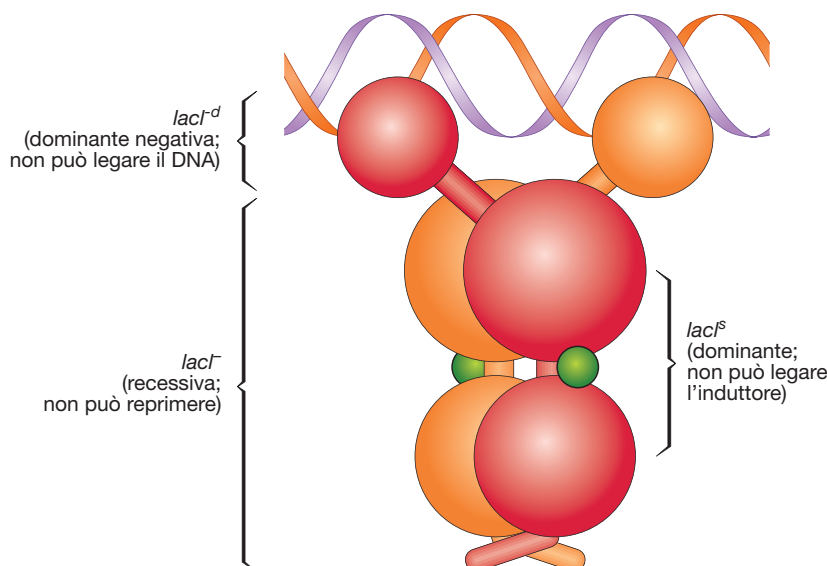
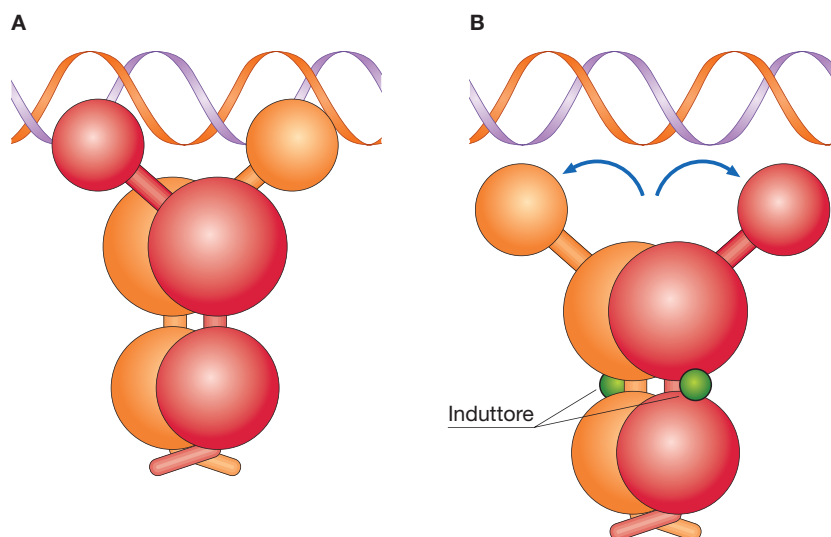


Figura 10.8 Mutazioni che alterano l'attività del repressore. Sono qui indicati tre tipi di mutazioni che alterano l'attività del repressore. Le mutazioni $lacI^-$ recessive, che non reprimono, possono mappare ovunque nella struttura della proteina. Le mutazioni $lacI^{-d}$ dominanti negative non possono reprimere perché è mutata la regione di legame con il DNA. Le mutazioni $lacI^S$, in generale dominanti, non possono portare all'induzione della trascrizione perché è mutato il sito di legame con l'induttore.

Figura 10.9 L'induttore (allolattosio) altera la conformazione del repressore. (A) In un dimero di repressore legato a un sito operatore i due domini di legame si trovano nella giusta conformazione e alla giusta distanza per legarsi ai due emisiti nel solco maggiore della doppia elica. (B) Il legame dell'induttore (allolattosio) induce un cambiamento conformazionale che allontana tra loro i domini di legame al DNA, cosicché il repressore perde la sua affinità per il DNA.



però in un'interazione a **bassa affinità**. La struttura dimerica, invece, permette ai due monomeri di repressore di legarsi simultaneamente mediante le teste alla sequenza dell'operatore, come indicato in **figura 10.9A**. Le teste entrano in contatto a livello del solco maggiore inserendosi in due giri d'elica consecutivi, assicurando così una interazione ad **alta affinità** e di gran lunga maggiore rispetto all'interazione da parte di un solo monomero. In **figura 10.9B** è rappresentato il repressore legato all'induttore; questa interazione provoca una modificazione allosterica del repressore che comporta un allontanamento tra loro dei domini che legano il DNA. Il repressore non è così più in grado di rimanere legato al suo sito operatore e libera il promotore per il legame dell'RNA polimerasi.

Abbiamo fin qui parlato dell'interazione del repressore Lac dimerico con un sito operatore. In realtà è stato poi trovato che il repressore lega contemporaneamente, in forma tetramERICA, due siti operatori. Il repressore tetramericO è tenuto insieme dalla formazione di un fascio di 4 α -eliche, ciascuna all'estremità C-terminale di un monomero. Come illustrato schematicamente in **figura 10.10A**, l'operatore *lac* è costituito da tre sequenze denominate O_1 , O_2 e O_3 . Il repressore Lac tetramericO si lega a due di essi, avvicinandoli e formando un'ansa di DNA la cui lunghezza può essere da circa 90 a 400 pb a seconda della distanza tra gli operatori selezionati. Il sito primario di legame del repressore è O_1 , localizzato all'inizio del gene *lacZ*, ed è quello con affinità più alta per il repressore. Gli altri sono a più bassa affinità e definiti deboli. O_2 si trova a 410 pb e O_3 a 83 pb, rispettivamente, a valle e a monte di O_1 . Il repressore tetramericO legherebbe il sito primario O_1 e uno dei due siti a bassa affinità, O_2 e O_3 . Tale legame multiplo influenza l'efficienza di repressione. Infatti, è stato dimostrato che la delezione degli operatori O_2 e O_3 determina un notevole abbassamento nell'azione di repressione. Diverse sono le ipotesi a favore di un legame iniziale aspecifico del repressore con il DNA, che scivolerebbe poi sulla doppia elica alla ricerca dei siti più affini. La **figura 10.10B** mostra un'immagine strutturale di un tetrametro del repressore legato a due operatori.

Controllo positivo dell'operone *lac* e ruolo di CAP

Qualche anno dopo la proposta da parte di Jacob e Monod del modello di controllo negativo dell'operone *lac*, è stato scoperto che lo stesso operone è soggetto anche a un controllo di tipo positivo. Infatti, affinché la trascrizione dell'operone *lac* sia attivata, non solo è necessaria la presenza dell'induttore lattosio che provoca il distacco del repressore dall'operatore, ma è anche necessaria l'assenza dal

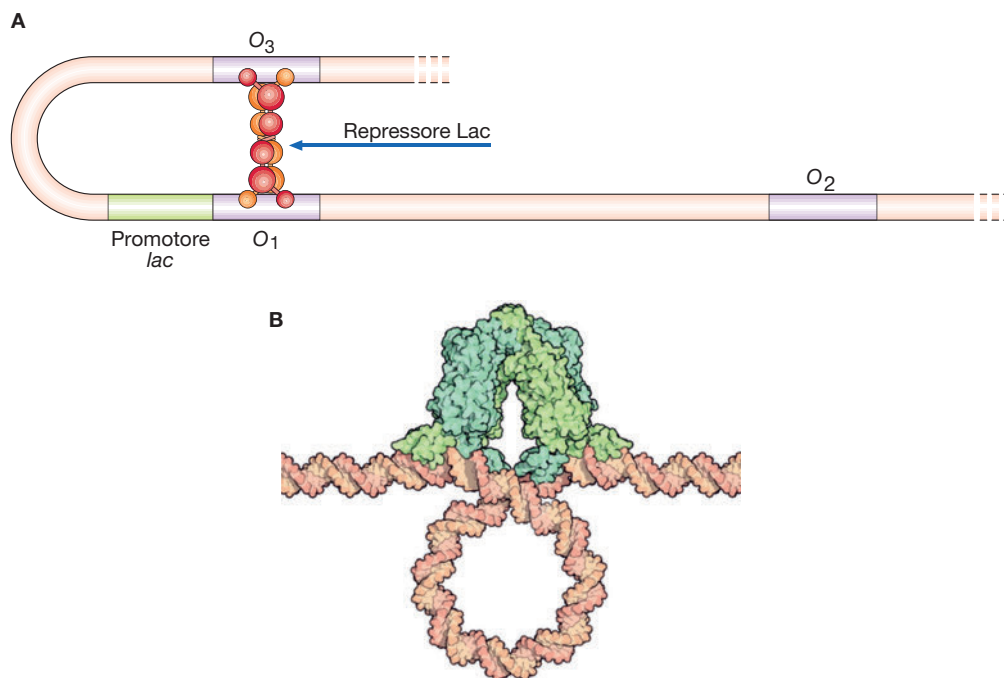


Figura 10.10 Il repressore Lac si lega contemporaneamente, in forma di tetramero, a due siti operatori. (A) Oltre al principale sito operatore chiamato O_1 , sono presenti due siti secondari: O_2 a valle e O_3 a monte. Due dimeri di repressore legano, rispettivamente, il sito principale e uno dei due siti secondari (O_3 in questa figura) e interagiscono tra loro mediante la regione C-terminale di

oligomerizzazione. Il legame del repressore tetramerico comporta l'acquisizione di una struttura ad ansa del DNA e risulta in una aumentata stabilità del legame repressore/operatore. (B) Struttura tridimensionale di un repressore tetramerico legato a due siti operatori. (Fonte: www.rcsb.org, ProteinDataBank.)

mezzo di glucosio, che, come vedremo, provoca il legame di un'altra proteina regolatrice la quale, legandosi a monte del promotore, ne stimola la trascrizione. Il significato di questa regolazione è che lo zucchero che *E. coli* utilizza più frequentemente è il glucosio, preferito come sorgente di energia. Solo in assenza di glucosio il batterio si adatta attivando uno o più operoni che gli permettono l'utilizzazione di altri zuccheri (per es. il lattosio che il batterio scinde producendo il glucosio che manca). Tale regolazione prende il nome di **repressione da catabolita** ed è controllata dal livello di cAMP presente nella cellula.

La **figura 10.11** illustra schematicamente questo meccanismo di controllo positivo dell'operone *lac*. In un punto del genoma distante dall'operone *lac* è localizzato il gene che codifica per la proteina regolatrice **CAP** (attivatrice dei geni catabolici), anche denominata **CRP** (recettore del cAMP: adenosina monofosfato 3'-5' ciclico). Invece, adiacente all'operone *lac*, subito a monte della regione promotore/operatore di cui abbiamo parlato nel paragrafo precedente, è localizzato il sito CAP, cioè una sequenza riconosciuta e legata dalla proteina regolatrice CAP solo quando questa è legata al cAMP. Il cAMP (**fig. 10.12B**) è sintetizzato, utilizzando come substrato l'ATP, dall'enzima adenilato ciclasi che si attiva solo se i livelli cellulari di glucosio sono molto bassi ed è quindi necessario attivare la trascrizione dell'operone *lac* per permettere la produzione di glucosio per scissione del lattosio. La struttura della proteina CAP legata al DNA è mostrata in **figura 10.12**. Questa proteina è un dimerico formato da 2 subunità di 22,5 kDa ciascuna, ed è costituita da un dominio per il riconoscimento della sequenza bersaglio sul DNA e un dominio di attivazione che, interagendo con l'RNA polimerasi, attiva la trascrizione. Come indicato nella **figura 10.11A**, CAP interagisce stabilmente con il suo sito bersaglio sul DNA solo se legata al cAMP. Questo legame di CAP/cAMP al DNA contribuisce ad aumentare la capacità dell'RNA polimerasi di trascrivere l'operone

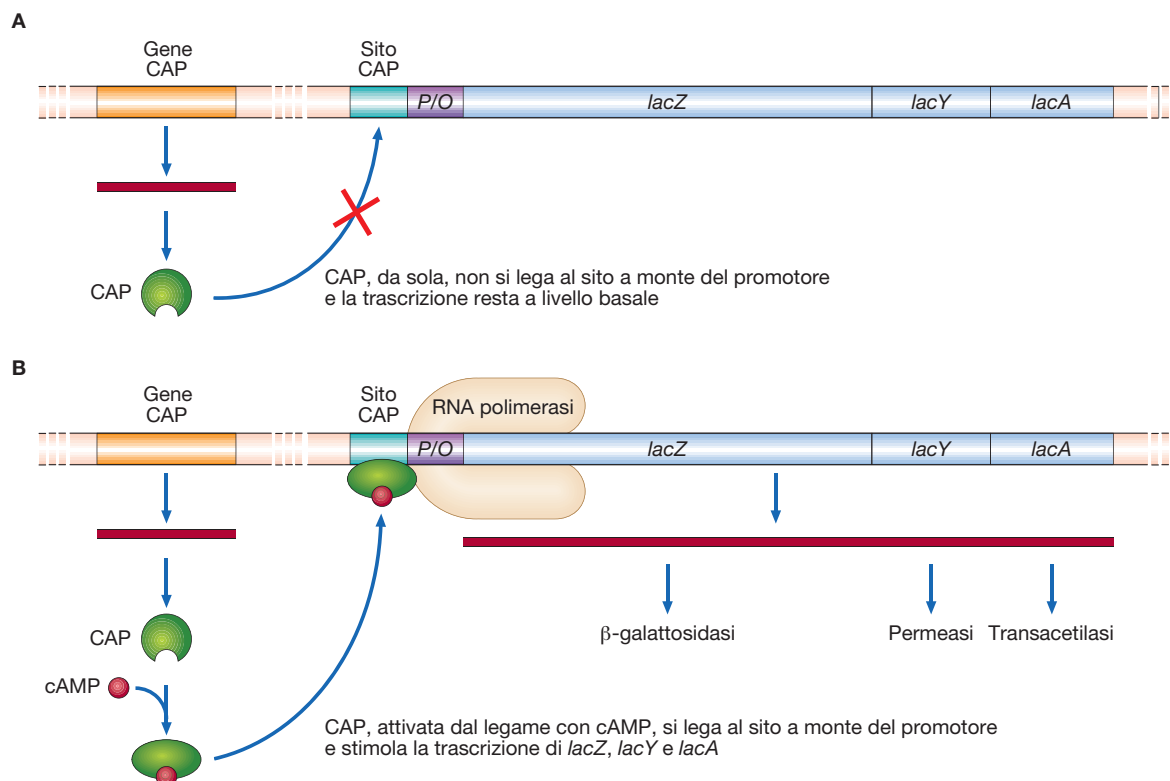
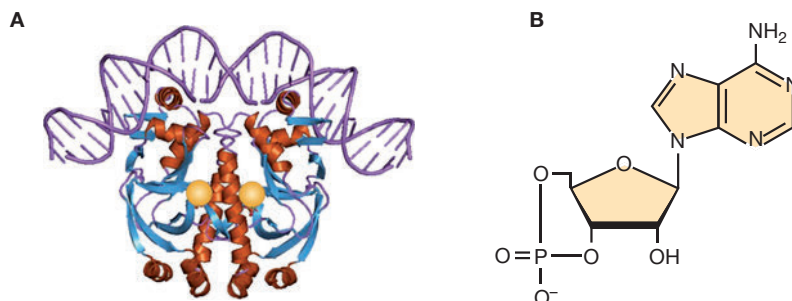


Figura 10.11 Controllo positivo dell'operone lattosio. Oltre al controllo negativo sopra descritto, l'operone *lac* è soggetto anche a un controllo di tipo positivo. (A) A monte della regione regolativa promotore/operatore dell'operone *lac* è presente un sito di legame per la proteina regolatrice CAP, prodotto da un gene loca-

lizzato a distanza nel genoma. CAP, da sola, non è in grado di riconoscere e legare il corrispondente sito bersaglio, determinando un basso livello di trascrizione. (B) In assenza di glucosio si ha un aumento del cAMP, che, legando CAP, la rende in grado di legare il sito CAP e di attivare così la trascrizione dei geni strutturali.

Figura 10.12 Struttura della proteina CAP e del suo ligando cAMP. (A) La proteina CAP, in forma dimerica, riconosce e lega il DNA mediante un motivo elica-giro-elica che interagisce con il solco maggiore del DNA, provocando una piegatura della doppia elica. In rosso sono indicate le regioni a struttura α -elica e in blu i foglietti β ; le sferette arancioni rappresentano il ligando cAMP. (B) Struttura del cAMP.



lac, che di per sé è provvisto di un promotore debole, cioè a bassa efficienza. A livello della sequenza del DNA viene indotta una modificazione conformazionale in quanto CAP, legandosi nel solco maggiore del DNA attraverso il motivo elica-giro-elica (HTH, *Helix-Turn-Helix*), provoca un piegamento della doppia elica di circa 90°. Questo ripiegamento facilita l'interazione diretta di CAP con la parte C-terminale della subunità α dell'RNA polimerasi.

L'analisi di mutanti di CAP che sono in grado di legare il DNA, ma non di attivare la trascrizione, ha permesso l'identificazione del dominio che interagisce direttamente con la polimerasi. CAP attiva la trascrizione reclutando la polimerasi su una serie di promotori privi dell'elemento UP, dove in genere interagisce l' α -CTD (dominio C-terminale della subunità α della polimerasi), ma che, come illustrato in **figura 10.13**, sono dotati del sito CAP. Il fatto che la funzione degli attivatori sia

essenzialmente quella di reclutare in modo stabile l'RNA polimerasi sul promotore è stato dimostrato con due eleganti esperimenti di "by-pass" dell'attivatore (**fig. 10.14**). Nel primo esperimento sono state inserite, a monte del gene *lacZ*, delle sequenze di legame specifiche per un attivatore artificiale. Questa molecola interagisce specificamente con l' α -CTD (dominio carbossi-terminale della subunità α della polimerasi), modificato proprio per far avvenire l'interazione. In questo caso la proteina artificiale è in grado di attivare con grande efficienza la trascrizione. In un altro esperimento, l' α -CTD è stato sostituito con la proteina CAP, e anche in questo caso la polimerasi modificata trascrive con grande efficienza. Questi due esperimenti dimostrano che per una trascrizione efficiente si deve formare un complesso stabile della polimerasi sul promotore e che gli attivatori sono responsabili di questo reclutamento. Poiché questa attivazione può essere realizzata da proteine diverse, si può dedurre che l'attivazione della trascrizione non coinvolge modificazioni allosteriche della polimerasi da parte degli attivatori.

CAP regola in maniera analoga l'attività di numerosi altri operoni, in particolare gli operoni per la sintesi di enzimi necessari all'utilizzo di altri zuccheri come fonte di energia, i cui promotori hanno come caratteristica comune l'assenza dell'elemento UP e la presenza di un sito CAP.

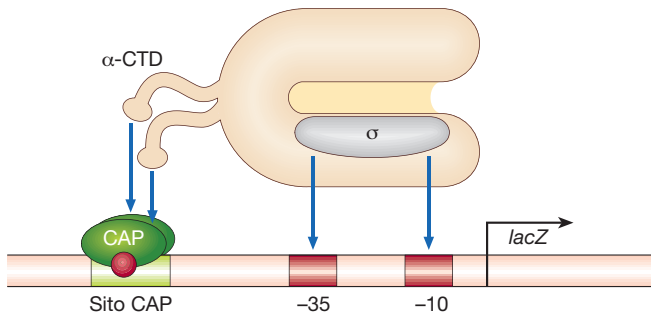


Figura 10.13 CAP lega direttamente la polimerasi. La proteina CAP attiva la trascrizione legandosi specificamente, come dimero, al suo sito di riconoscimento che si trova a monte del promotore e interagendo direttamente con i domini C-terminali delle due subunità α della polimerasi (α -CTD). CAP attiva soprattutto geni privi dell'elemento UP a cui in genere si legano gli α -CTD.

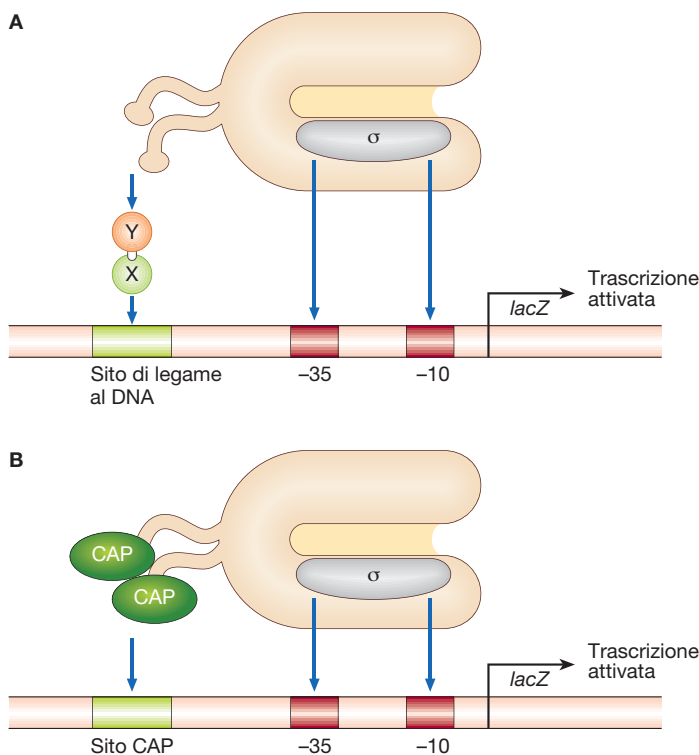


Figura 10.14 Esperimenti di bypass dell'attivatore. (A) A monte del promotore del gene *lacZ* sono state inserite sequenze di legame specifiche per un attivatore artificiale costituito da due domini: un dominio Y (arancione) che si lega specificamente al sito di legame inserito nel DNA a monte del promotore. La trascrizione è attivata con grande efficienza. (B) La proteina CAP è stata fusa direttamente alla polimerasi, eliminando le due estremità C-terminali della subunità α . Anche in questo caso la trascrizione è fortemente attivata, dimostrando che sono le interazioni dirette dell'attivatore con la polimerasi che la reclutano sul promotore attivando così la trascrizione.

Interazione della regione promotore/operatore con CAP, repressore Lac e RNA polimerasi

Nei paragrafi precedenti i due meccanismi di controllo, uno negativo e l'altro positivo, che regolano la trascrizione dei geni strutturali dall'operone *lac* sono stati descritti separatamente l'uno dall'altro. In realtà, i due meccanismi si sovrappongono sia funzionalmente che fisicamente, in quanto i componenti in gioco – RNA polimerasi, repressore Lac e CAP – interagiscono tutti e tre nello spazio limitato di un centinaio di pb. Il motivo strutturale alla base del legame di ambedue le proteine regolatrici al DNA è il dominio elica-giro-elica (*helix-turn-helix*) che abbiamo menzionato parlando del repressore e che approfondiremo nel prossimo capitolo (*vedi* par. 11.7 e fig. 11.20). La **figura 10.15** riporta la sequenza nucleotidica dei siti di riconoscimento per le due proteine regolatrici e per l'RNA polimerasi. Il sito di legame per CAP (sito CAP) è di circa 20 pb ed è situato a monte del promotore *lac*, mentre O_1 , principale sito di legame per il repressore, è localizzato a partire dalle ultime basi che sono anche riconosciute dall'RNA polimerasi. È facile intuire che il repressore agisce occupando fisicamente il sito e impedendone la trascrizione. In particolare, l'RNA polimerasi lega debolmente il promotore in assenza della proteina attivatrice e questo si verifica perché la sequenza -35 (**fig. 10.15**) non è ottimale per il legame e il promotore manca dell'elemento UP. CAP si lega a una sequenza simile in lunghezza all'operatore *lac*, che si trova 60 pb a monte del sito d'inizio della trascrizione. La proteina attivatrice, legandosi come dimerico alle sequenze di riconoscimento, recluta stabilmente l'RNA polimerasi sul promotore. Tale meccanismo cooperativo stabilizza e ottimizza la formazione del complesso aperto e, quindi, il livello generale di trascrizione. Le subunità dell'RNA polimerasi che permettono tale attività sono indicate in **figura 10.13**. Da una parte l' α -CTD, strutturalmente legato al dominio N-terminale (NTD) della subunità α mediante una regione linker flessibile, protrude dalla struttura enzimatica e può così interagire con CAP, che è legata al sito CAP a monte del promotore. D'altro canto il fattore σ^{70} riconosce e lega la sequenza nucleotidica a livello degli elementi -35 e -10 .

Oltre ai meccanismi esposti finora, ve ne sono altri che intervengono nel delicato processo della regolazione della trascrizione. Mentre la gran parte degli attivatori funziona, come CAP per l'operone *lac*, reclutando l'RNA polimerasi sul promotore, vi sono casi in cui il funzionamento dell'attivatore è basato su un meccanismo allosterico. Descriviamo qui, come esempio, il caso di NtrC. NtrC controlla l'espressione dei geni *glnA* coinvolti nel metabolismo dell'azoto e media un cambiamento allosterico dell'RNA polimerasi, da un complesso d'inizio chiuso verso quello aperto (**fig. 10.16**). Infatti, in assenza di segnali intracellulari e, conseguentemente, di NtrC attivo, l'RNA polimerasi che per la trascrizione di

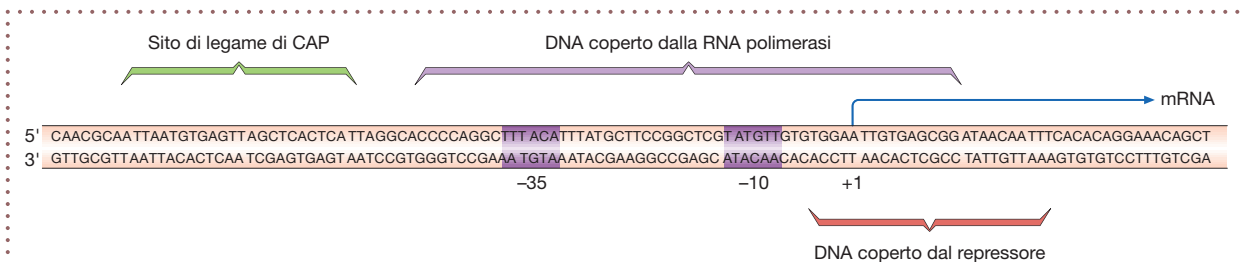


Figura 10.15 Sequenza della regione del promotore *lac*. La regione regolativa a monte dell'operone *lac* comprende gli elementi di legame per l'RNA polimerasi (promotore), per il repressore (operatore) e per la proteina CAP (sito CAP). Il sito operatore a cui si lega

il repressore si trova a valle del promotore (sequenze -10 e -35) e parzialmente sovrapposto ad esso, mentre il sito CAP è posizionato a monte del promotore, in modo tale da richiamare l'RNA polimerasi sul promotore.

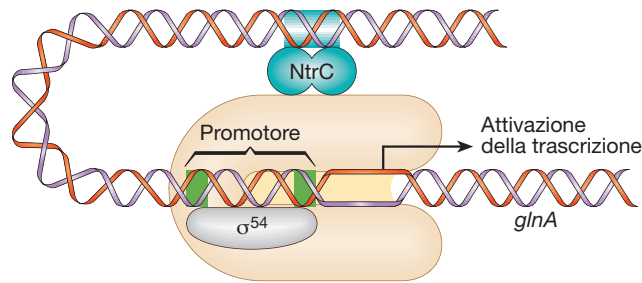


Figura 10.16 Regolazione positiva da parte di NtrC dei geni per il metabolismo dell'azoto. NtrC, come la proteina CAP, ha un sito di legame al DNA e un sito di attivazione. Questa proteina viene fosforilata quando nella cellula ci sono bassi livelli di azoto e, solo allora, è in grado di legarsi al DNA. NtrC legato ai suoi siti interagisce anche

con l'RNA polimerasi legata al promotore, inducendo la formazione di un'ansa sul DNA. Questa polimerasi ha una subunità σ diversa (σ^{54}) che non è in grado di indurre la formazione del complesso aperto. L'attività ATPasica del complesso NtrC induce un cambiamento conformazionale in σ^{54} che favorisce la formazione del complesso aperto.

glnA utilizza σ^{54} , resta legata al promotore in conformazione chiusa. In presenza di un segnale specifico, ovvero basse concentrazioni di azoto intracellulari, NtrC viene attivato mediante fosforilazione e, come mostrato in figura, può legarsi sotto forma di dimeri 150 pb a monte del promotore e, contemporaneamente, alla RNA polimerasi mediante il dominio di riconoscimento per σ^{54} . Queste due interazioni implicano la formazione di un'ansa di DNA. Con la sua attività ATPasica, NtrC fornisce l'energia per la reazione di apertura del complesso chiuso polimerasi/promotore e attiva la trascrizione.

10.3 Operone del triptofano

Riportiamo ora un esempio di operone reprimibile mediante un meccanismo di controllo negativo. In *E. coli* l'**operone *trp* (triptofano)** è costituito da un gruppo di geni contigui che codificano per una serie di enzimi responsabili della sintesi dell'amminoacido. Come tutti gli operoni implicati nella produzione degli enzimi delle vie metaboliche per la sintesi degli amminoacidi, si tratta di un sistema reprimibile, cioè l'espressione di questi geni viene repressa dall'aggiunta nel mezzo di coltura dell'amminoacido corrispondente. L'operone triptofano è soggetto a due meccanismi di regolazione che operano a livelli distinti: uno a livello d'inizio, l'altro a livello di terminazione della trascrizione. Parleremo ora del primo meccanismo, che consiste in un controllo di tipo negativo di un sistema reprimibile (**finestra 10.2, fig. 10F2.1**). Parleremo invece, più avanti, del secondo meccanismo, chiamato "attenuazione", che opera modulando un evento di terminazione della trascrizione. In **figura 10.17** è rappresentata la struttura dell'operone con i suoi geni strutturali: *trpE* (codificante per l'enzima antranilato sintetasi), *trpD* (per una fosforibosil-antranilato transferasi), *trpC* (per una fosforibosil-antranilato indol-isomerasi glicerol fosfato sintetasi), *trpB* (per una triptofano β -sintetasi), *trpA* (per la triptofano α -sintetasi). Come per l'operone *lac*, la regolazione dell'operone *trp* è di tipo negativo e si realizza tramite un repressore. Questo però, a differenza del repressore Lac, è di per sé inattivo e viene attivato dal legame di una molecola di triptofano che agisce perciò da **corepressore**. Infatti, in presenza di triptofano intracellulare questo si lega al repressore inducendo un cambiamento conformazionale che lo rende capace di legare l'operatore e bloccare la trascrizione (**fig. 10.17A**). Al contrario, a basse concentrazioni di triptofano intracellulare, il repressore non è in grado di legare l'operatore e l'RNA polimerasi può iniziare la sintesi dell'mRNA (**fig. 10.17B**). La **finestra 10.2** riassume schematicamente e mette a confronto i possibili diversi circuiti regolativi che nei procarioti controllano, con meccanismi di regolazione positiva o negativa, sia gli operoni inducibili che quelli reprimibili.

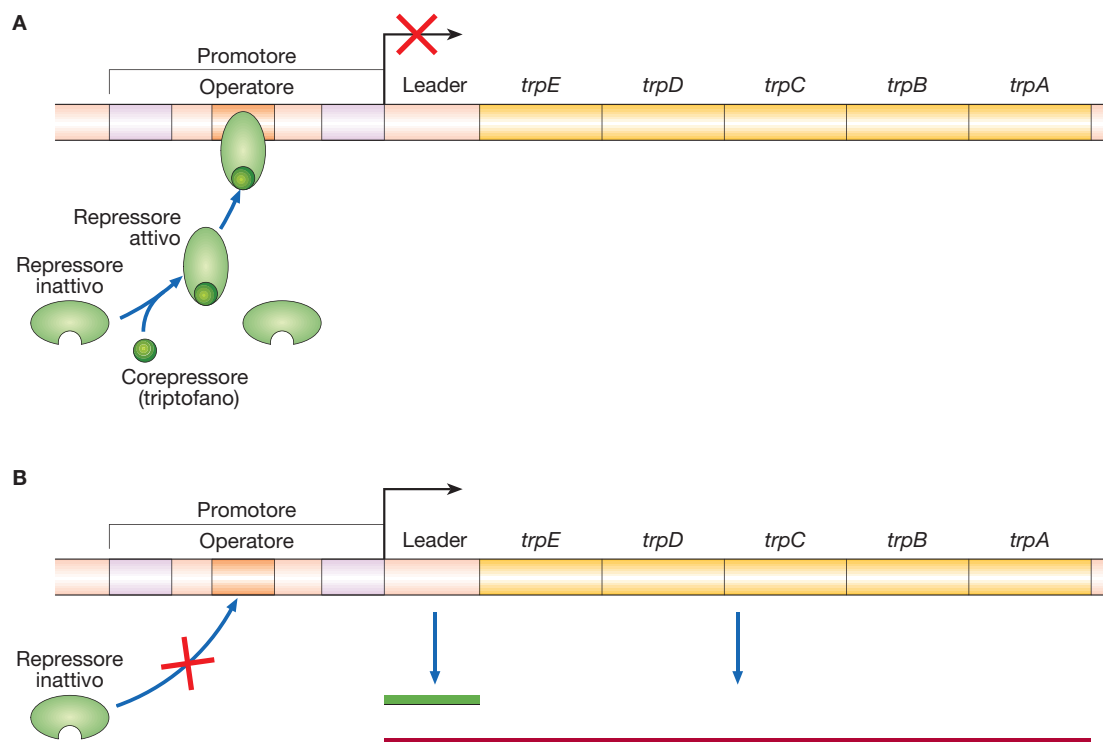


Figura 10.17 L'operone triptofano. (A) I geni strutturali per gli enzimi implicati nel metabolismo del triptofano sono disposti a valle della regione promotore/operatore. La trascrizione è regolata da un repressore, di per sé inattivo, che viene attivato dal legame del triptofano (corepressore). (B) In carenza di triptofano

la trascrizione viene attivata con produzione di due trascritti alternativi: il trascritto attenuato che comprende solo la porzione iniziale dell'operone e l'mRNA policistronico completo. Questo meccanismo di controllo, chiamato "attenuazione", è descritto nelle prossime figure.

FINESTRA 10.2

OPERONI INDUCIBILI E REPRIMIBILI, CONTROLLI NEGATIVI E CONTROLLI POSITIVI

Come abbiamo visto nella finestra 10.1, Jacob e Monod, sulla base dello studio dell'operone *lac*, hanno proposto che la regolazione sia sostanzialmente dovuta a elementi agenti in *trans*, quali fattori di trascrizione e proteine regolatrici, che interagiscono con elementi agenti in *cis*, quali il promotore per l'RNA polimerasi e l'operatore per le proteine regolatrici. Questo concetto di base è estendibile a molti altri operoni con numerose varianti: operoni inducibili, reprimibili e con meccanismi di controllo negativi o positivi.

Per **operoni inducibili** si intendono quelli la cui trascrizione viene indotta dall'aggiunta al mezzo di coltura di una molecola segnale (induttore), mentre **operoni reprimibili** sono quelli che vengono repressi dalla molecola segnale (corepressore).

L'esempio tipico di operone inducibile è l'operone *lac* che abbiamo appena descritto, in cui l'induttore è rappresentato dal lattosio. Molti altri operoni sono inducibili; in particolare tutti gli operoni che codificano per gli enzimi delle varie vie metaboliche necessarie all'utilizzazione di altri zuccheri. Per esempio, l'arabinosio induce l'operone *ara* costituito da 3 geni che codificano per le proteine necessarie alla sua utilizzazione e così via per altri zuccheri.

Sono invece reprimibili quegli operoni la cui trascrizione viene repressa, anziché indotta, dalla molecola segnale. Per esempio, *E. coli* ha i geni che codificano per gli enzimi delle vie metaboliche richieste per la sintesi dei vari amminoacidi; tuttavia, se un amminoacido è presente nel mezzo di coltura, esso causa il blocco dell'espressione del corrispondente operone, con notevole risparmio di energia per la cellula. Tutti gli operoni per la sintesi degli amminoacidi, e anche quelli per la sintesi di altre molecole necessarie alla cellula, sono di tipo reprimibile. Come esempio descriveremo più avanti il caso tipico dell'operone *trp* per la produzione degli enzimi necessari alla sintesi dell'amminoacido triptofano.

La regolazione di tutti questi operoni, inducibili o reprimibili, può basarsi su meccanismi che possono essere di due tipi: (1) meccanismi di controllo negativo e (2) meccanismi di controllo positivo. Come abbiamo già detto, la regolazione negativa è quella in cui la proteina *trans*-agente prodotta dal gene regolatore si lega a sequenze *cis*-agenti, sovrapposte o vicine al promotore, e blocca la trascrizione; il distacco della proteina permette l'inizio della trascrizione. Esempio tipico di controllo negativo è quello esercitato dal repressore Lac. Nella regolazione positiva, invece, la proteina *trans*-agente prodotta

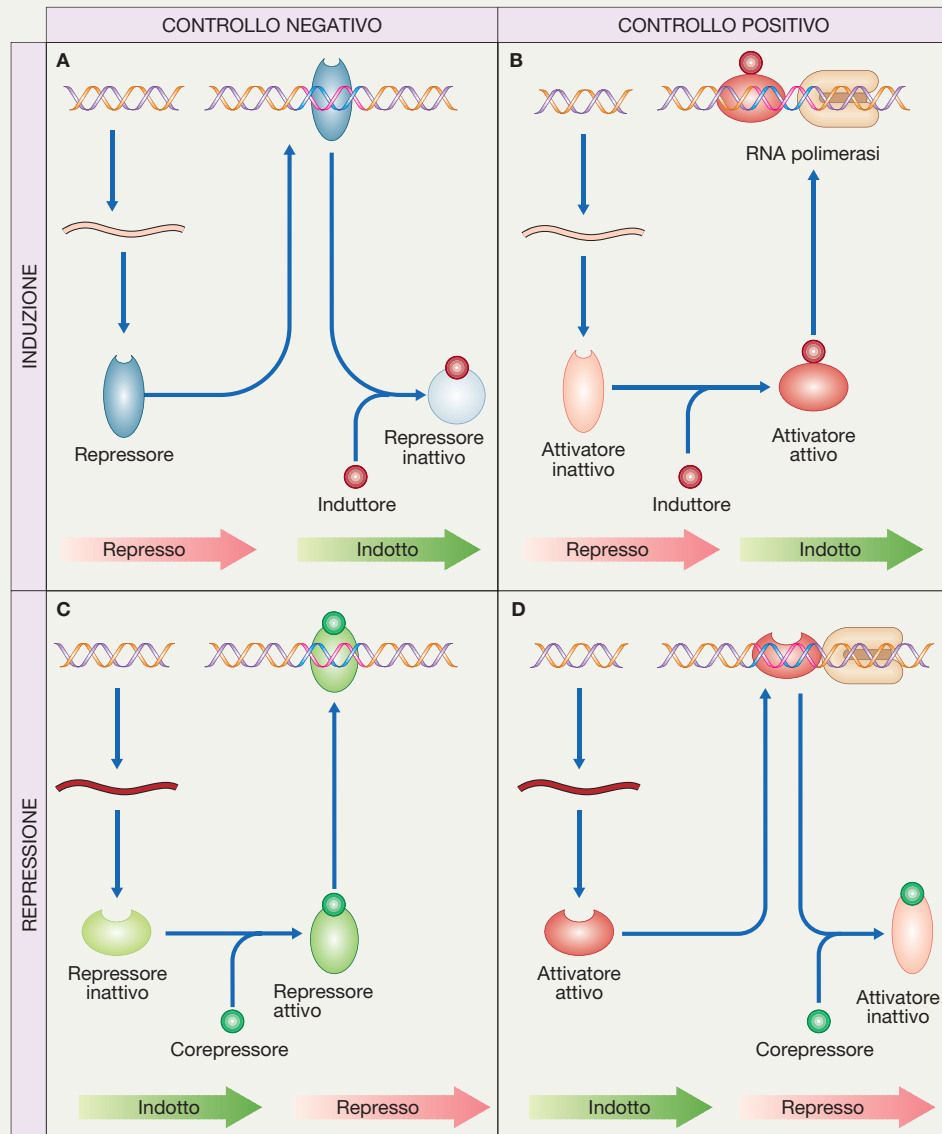


Figura 10F2.1 Il controllo positivo e negativo può regolare l'induzione e la repressione. La figura illustra le varie combinazioni di induzione e repressione realizzate dal controllo positivo e negativo. Il controllo negativo di un sistema inducibile (riquadro A) comporta la repressione da parte di un repressore attivo; si ha induzione quando il repressore viene disattivato dal legame dell'induttore. Il controllo positivo di un sistema inducibile (riquadro B) risulta, in assenza di induttore,

in uno stato represso perché l'attivatore di per sé non funziona; si ha induzione quando l'interazione con l'induttore lo attiva. Nel caso di repressione mediante controllo negativo (riquadro C) il repressore non è attivo di per sé, ma viene attivato dall'interazione con il corepressore. In ultimo (quadro D) la repressione mediante controllo positivo è dovuta a un attivatore che viene inattivato dall'interazione con il corepressore. (Fonte: ripresa e rielaborata da B. Lewin, *Genes VI*.)

dal gene regolatore si lega a sequenze *cis*-agenti, nelle vicinanze del promotore, ma alle volte anche a una certa distanza, e in qualche modo stimola la trascrizione. Così, per esempio, l'operone inducibile *ara* (arabinosio) è soggetto a un controllo di tipo positivo. Analogamente, come vedremo poco oltre, anche la regolazione esercitata dalla proteina CAP sulla trascrizione dello stesso operone *lac* è di tipo positivo; quindi l'operone *lac* è soggetto a due meccanismi di regolazione: quello negativo, descritto originalmente da Jacob e Monod, e quello positivo. Come schematizzato in **figura 10F2.1**, per un operone, il fatto di essere inducibile o reprimibile, e a controllo positivo o negativo, sono due aspetti diversi e indipendenti, potendo

esistere le diverse combinazioni: (1) operoni inducibili a controllo negativo, (2) inducibili a controllo positivo, (3) reprimibili a controllo negativo, (4) reprimibili a controllo positivo. Abbiamo già descritto il caso di operone inducibile a controllo negativo con l'esempio classico della regolazione da parte del repressore dell'operone *lac*. Descriveremo appresso la regolazione inducibile a controllo positivo dello stesso operone *lac* da parte della proteina regolatrice CAP e riporteremo poi l'operone *trp* (triptofano) come esempio di operone reprimibile a controllo negativo. Esempi di operoni reprimibili a controllo positivo, benché possibili e forse presenti in natura, non sono ancora stati descritti.

Regolazioni a livello di terminazione della trascrizione

La maggior parte del controllo della trascrizione nei procarioti agisce sulla fase d'inizio della trascrizione. Esistono però alcuni casi in cui l'attività genica è regolata a livello di terminazione. Infatti, la modulazione dell'efficienza di un terminatore della trascrizione ha un effetto sulla trascrizione delle sequenze nucleotidiche a valle. Ci sono due principali processi di questo tipo: l'antiterminazione e l'attenuazione.

L'**antiterminazione**, che è spesso implicata nell'attivazione a cascata dei geni precoci, intermedi e tardivi dei fagi, si basa sull'inibizione della terminazione da parte di proteine regolatrici, con conseguente trascrizione dei geni adiacenti a valle. Parleremo di questo meccanismo in dettaglio nel prossimo paragrafo sulla regolazione del fago lambda.

Descriviamo qui, invece, il fenomeno dell'**attenuazione**, prendendo in considerazione come esempio l'operone *trp*. Abbiamo visto sopra che l'operone *trp* reprimibile è regolato a livello d'inizio della trascrizione da un meccanismo di tipo negativo responsabile della regolazione principale connessa alla presenza o assenza di triptofano nel mezzo di crescita. Lo stesso operone è soggetto a un secondo livello di regolazione, più fine, basato su un meccanismo che prende, appunto, il nome di attenuazione. Questo meccanismo interrompe prematuramente la sintesi dell'mRNA in risposta alla presenza di triptofano quando questo, sebbene a bassi livelli, è sufficiente per le attività metaboliche cellulari (**fig. 10.17B**). Infatti, in presenza di basse concentrazioni di triptofano, l'RNA polimerasi che ha iniziato la trascrizione non sintetizza l'intero mRNA policistronico, ma si interrompe a un sito specifico dopo solo circa 140 nucleotidi. Quando invece il triptofano è limitante, essendo presente a concentrazioni ancora più basse, la polimerasi non si ferma, ma procede nella trascrizione dell'intero operone portando alla produzione degli enzimi per la sintesi dell'amminoacido.

Il meccanismo di questa regolazione è spiegabile in termini di struttura di una sequenza nucleotidica che si trova compresa tra il sito d'inizio della trascrizione e la fase di lettura aperta (ORF) del primo gene strutturale, *trpE*. Questa sequenza, chiamata **sequenza leader**, è mostrata come mRNA in **figura 10.18**; essa è costituita da 161 nucleotidi e contiene una breve ORF, cioè un codone di inizio seguito da 13 codoni per amminoacidi, tra cui due codoni per triptofano adiacenti, e un codone di terminazione. Come mostrato in figura, sovrapposte e a valle di questa breve sequenza codificante sono presenti sequenze complementari (1-2-3-4) che possono appaiarsi a formare strutture alternative (**fig. 10.19**). In particolare, la sequenza 3 si può appaiare sia con la 4 (**fig. 10.19A**) che con la 2 (**fig. 10.19C**). Nel primo caso, la forcina che si forma è un tipico terminatore di trascrizione intrinseco o rho-indipendente; infatti, subito dopo la forcina si trova una sequenza di 7 U (**fig. 10.18**). Invece, la forcina che si forma nella struttura alternativa, in cui 3 è appaiata con 2, non è un terminatore della trascrizione.

Come funziona la regolazione per attenuazione? In presenza di basse ma sufficienti concentrazioni intracellulari di triptofano, e conseguentemente dei corrispondenti tRNA carichi (Trp-tRNA^{Trp}) (*vedi* cap. 14), i ribosomi leggono l'ORF presente nella sequenza leader producendo un peptide di 13 amminoacidi, chiamato **peptide leader**, che verrà poi degradato. Nel fare questo i ribosomi impediscono l'appaiamento della sequenza 2 con la sequenza 3, la quale si appaia con la 4 formando la forcina (3/4). Questa, che come abbiamo detto è un terminatore intrinseco, segnala all'RNA polimerasi la terminazione della trascrizione a monte dell'inizio delle sequenze codificanti con produzione di un breve mRNA chiamato **attenuato**. L'mRNA attenuato e il peptide leader andranno incontro a degradazione. Al contrario, a concentrazioni di triptofano ancora più basse e insufficienti, i ribosomi che leggono la breve ORF vanno in stallo, cioè si

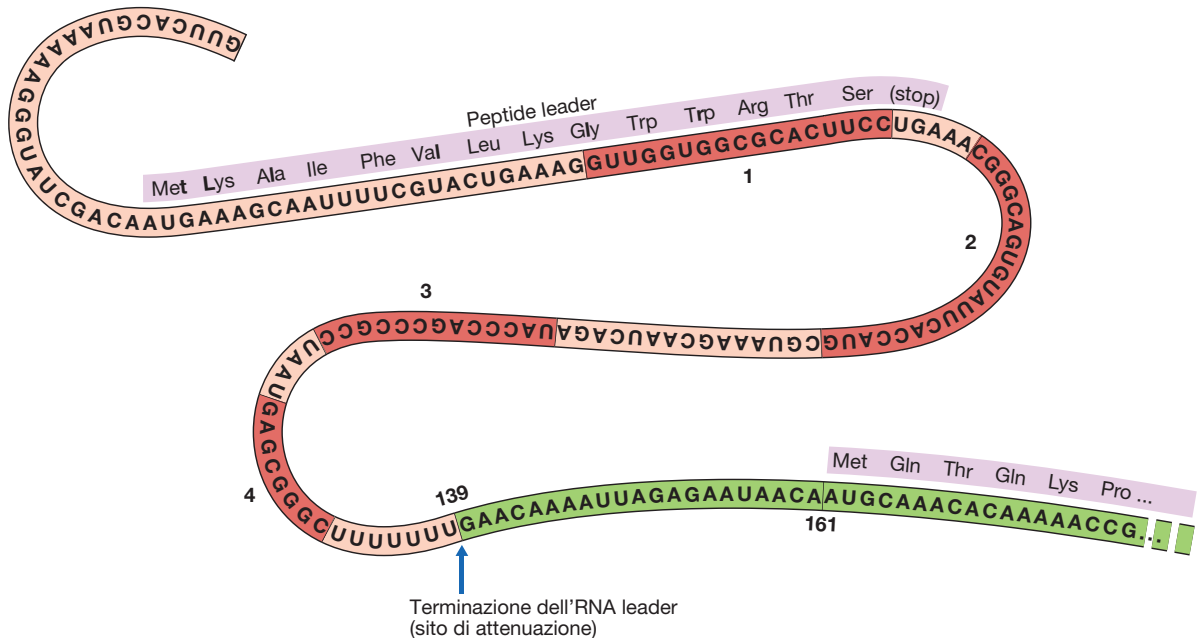


Figura 10.18 Struttura dell'attenuatore *trp*. La sequenza dell'mRNA che codifica per un breve peptide denominato "leader" è sovrapposta a quattro regioni definite 1, 2, 3 e 4, che una volta trascritte in RNA possono appaiarsi tra loro in due modi alternativi (vedi fig. 10.19). Si noti la serie di U caratteristica di

un terminatore della trascrizione. In viola è evidenziata, a monte, la sequenza degli amminoacidi del peptide leader dove sono presenti due residui di triptofano adiacenti e, più a valle, l'inizio della sequenza di amminoacidi della prima proteina codificata dall'operone.

fermano in corrispondenza dei due codoni per il triptofano, in quanto scarseggiano i $\text{Trp-tRNA}^{\text{Trp}}$. In questa situazione la regione 2 si può appaiare con la 3 formando la forcina alternativa (2/3) che non è un terminatore di trascrizione. L'RNA polimerasi procede, quindi, trascrivendo l'intero operone con tutti i suoi geni strutturali.

Il meccanismo di attenuazione non regola solo l'operone *trp*, ma anche gli operoni che codificano per gli enzimi delle vie metaboliche per la sintesi degli altri amminoacidi. Naturalmente, nel peptide leader di ciascun operone si trovano vari codoni per l'amminoacido corrispondente all'operone stesso. Per esempio, nell'operone *his* (istidina) l'ORF per il peptide leader di 15 amminoacidi contiene 6 codoni consecutivi per istidina, mentre l'ORF per il peptide leader di 21 amminoacidi dell'operone *thr* (treonina) ne contiene 8 per la treonina.

Va notato che il meccanismo di attenuazione può funzionare nei procarioti, dove trascrizione e traduzione sono processi accoppiati: infatti, i ribosomi legano l'mRNA nascente e lo traducono seguendo da vicino l'RNA polimerasi che lo sta sintetizzando. Questo fa sì che il processo di trascrizione possa essere influenzato dalla traduzione dell'mRNA stesso. Invece, nelle cellule eucariotiche, dove trascrizione e traduzione avvengono in compartimenti separati, rispettivamente nucleo e citoplasma, il meccanismo di attenuazione non può funzionare.

10.4 Controllo dell'espressione genica nel fago lambda

Come ricordato nel paragrafo 1.2, i batteriofagi sono stati adottati dal "gruppo del fago" a metà del secolo scorso per lo sviluppo della Genetica e Biologia Molecolare. I fagi rappresentano importanti sistemi modello per la comprensione dei meccanismi biologici di base, tra cui la trascrizione e la sua regolazione. Infatti,

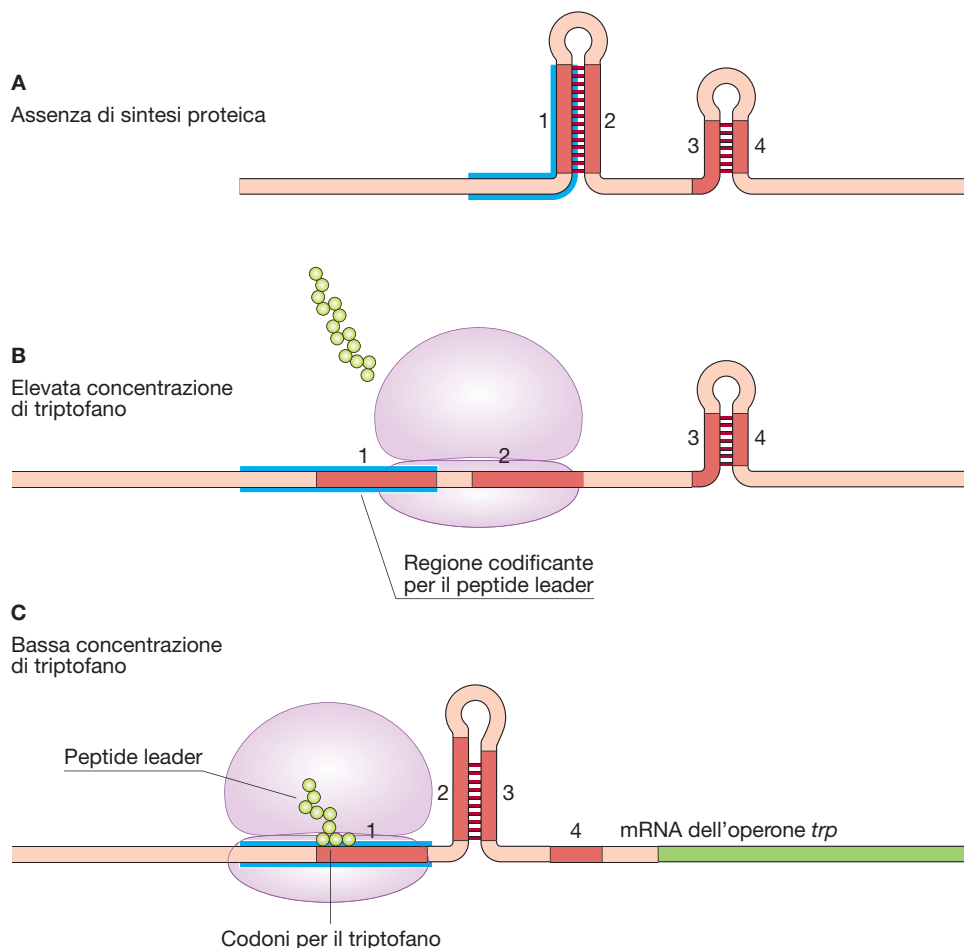


Figura 10.19 Meccanismo dell'attenuatore *trp*. (A) Le regioni 1, 2, 3 e 4 dell'RNA nascente possono appaiarsi a formare le forcine 1-2 e 3-4; quest'ultima è un'efficiente forcina di terminazione. (B) Quando nella cellula è presente molto triptofano, e non vi è quindi necessità di attivare l'operone, il ribosoma che traduce la sequenza leader dell'mRNA nascente coprirà le sequenze 1 e 2 impedendo la formazione della forcina 1-2 ma non della 3-4, che quindi pro-

vocherà interruzione della trascrizione a monte delle ORF dei geni strutturali. (C) In condizioni di bassa concentrazione di triptofano il ribosoma si arresterà sui due codoni *trp* presenti nella sequenza leader, coprendo la sequenza 1 ma non la 2; questa è libera di formare la forcina alternativa 2-3, energeticamente favorita a discapito della 3-4. Verrà così a mancare il terminatore trascrizionale e la trascrizione potrà proseguire per tutto l'operone policistronico.

tutti i fagi, più o meno grandi e complessi, esprimono i loro geni, durante il loro breve ciclo vitale, con una precisa sequenza temporale del tipo: geni precoci, geni intermedi e geni tardivi. Abbiamo visto nel paragrafo 9.4 che, in casi relativamente semplici come il fago SPO1 di *Bacillus subtilis*, ciò è dovuto a una cascata di attivazioni della trascrizione da parte di fattori sigma alternativi. In confronto a SPO1, il fago lambda (λ) è un sistema notevolmente più complesso, sia per il numero di geni, circa 50, sia soprattutto per il fatto che esso può seguire due vie alternative per perpetuare il proprio genoma: (1) il **ciclo litico**, comune a tutti i batteriofagi, in cui il virus infetta la cellula entro la quale si moltiplica in pochi minuti e che culmina con la distruzione della cellula ospite e il rilascio della progenie fagica; (2) il **ciclo lisogenico** in cui, come per alcuni altri fagi detti "temperati", il genoma fagico viene integrato nel cromosoma della cellula ospite e propagato con questo alle cellule figlie; solo particolari condizioni di "induzione" possono indurre l'escissione del genoma del fago da quello cellulare e la ripresa del ciclo litico. Questo comportamento implica fini processi regolativi dell'attività genica che, come vedremo, utilizzano meccanismi di controllo positivi e negativi che si intersecano e che operano a livello sia dell'inizio sia della termina-



Figura 10.20 Mark Ptashne. Nato a Chicago nel 1940, è ora al Memorial Sloan-Kettering Cancer Center, New York. Ptashne ha dato un fondamentale contributo alle conoscenze della regolazione della trascrizione, soprattutto chiarendo i meccanismi molecolari che regolano il ciclo litico e il ciclo lisogenico del fago lambda. Nel 1986 ha descritto queste ricerche e un modello di interpretazione in un libro che è diventato

un classico sia per il contenuto che per la forma chiara e precisa di trattazione dell'argomento (prima edizione americana: *A genetic switch*, CSHL Press, 1986; prima edizione italiana condotta sulla terza edizione americana: *Regolazione genica*, Zanichelli, 2006). Questo nostro paragrafo sulla regolazione del fago lambda e alcune delle figure di questo testo sono ispirate al libro di Ptashne.

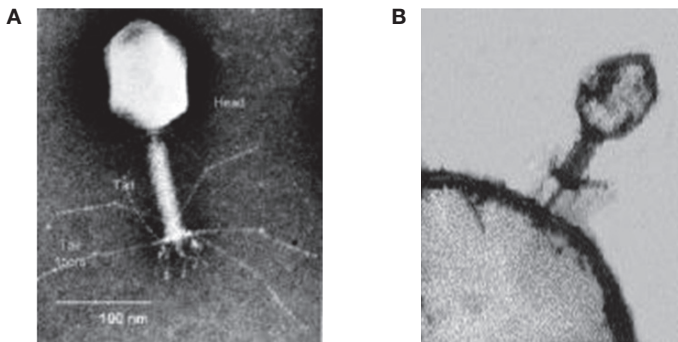


Figura 10.21 Il batteriofago lambda al microscopio elettronico. (A) Particelle di fago lambda mature; il diametro del capside è di circa 600 Å. (B) Una cellula di *Escherichia coli* sta per essere infettata dal fago.

zione della trascrizione. I risultati ottenuti negli approfonditi studi dei cicli litico e lisogenico del fago lambda hanno contribuito a definire concetti fondamentali generali validi anche per i sistemi eucariotici, dai *trans*-attivatori della trascrizione fino ai controlli post-trascrizionali ed epigenetici. Tra i numerosi ricercatori che hanno contribuito a questi studi, sicuramente lo statunitense Mark Ptashne (**fig. 10.20**) ha avuto un ruolo fondamentale.

La **figura 10.21A** mostra una fotografia al microscopio elettronico di particelle di fago mature dove è possibile vedere il capside e la coda che viene utilizzata per iniettare il DNA fagico nel batterio. La **figura 10.21B** mostra una cellula di *E. coli* che sta per essere infettata da un fago.

I due cicli vitali del fago lambda: ciclo litico e ciclo lisogenico

Quando un fago lambda infetta una cellula batterica, esso può scegliere di procedere con il ciclo litico o con quello lisogenico. La **figura 10.22** mostra i due possibili cicli del fago. La scelta tra il ciclo litico e il ciclo lisogenico è regolata in maniera molto fine da una serie di eventi e da proteine regolatrici che agiscono in *trans* su una particolare regione del genoma di lambda, detta **regione di controllo**.

In **figura 10.23** è riportata la mappa circolare del genoma del fago lambda, che è costituito da circa 50 kpb di DNA che codificano per circa 50 geni. Come abbiamo già detto nel paragrafo 4.1, all'interno della testa del fago il DNA è in

Figura 10.22 I due possibili cicli del fago lambda. Dopo l'infezione il fago può intraprendere due vie alternative: il *ciclo litico* (mostrato a destra) in cui il virus si replica producendo una numerosa progenie e che si conclude con la lisi della cellula e il rilascio dei nuovi fagi; il *ciclo lisogenico* (mostrato a sinistra) che comporta l'integrazione del genoma fagico nel cromosoma del batterio ospite e che è caratterizzato dalla presenza del repressore, proteina responsabile del mantenimento del ciclo lisogenico stesso.

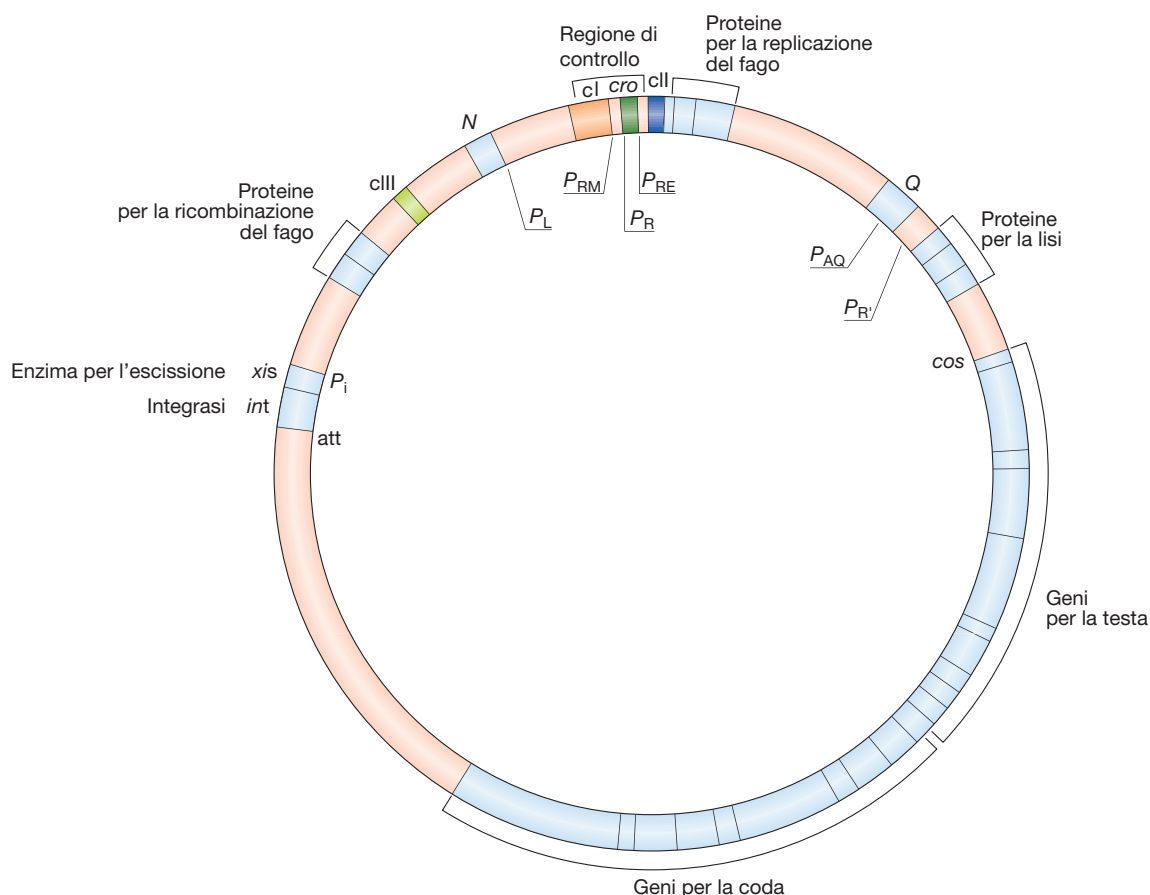
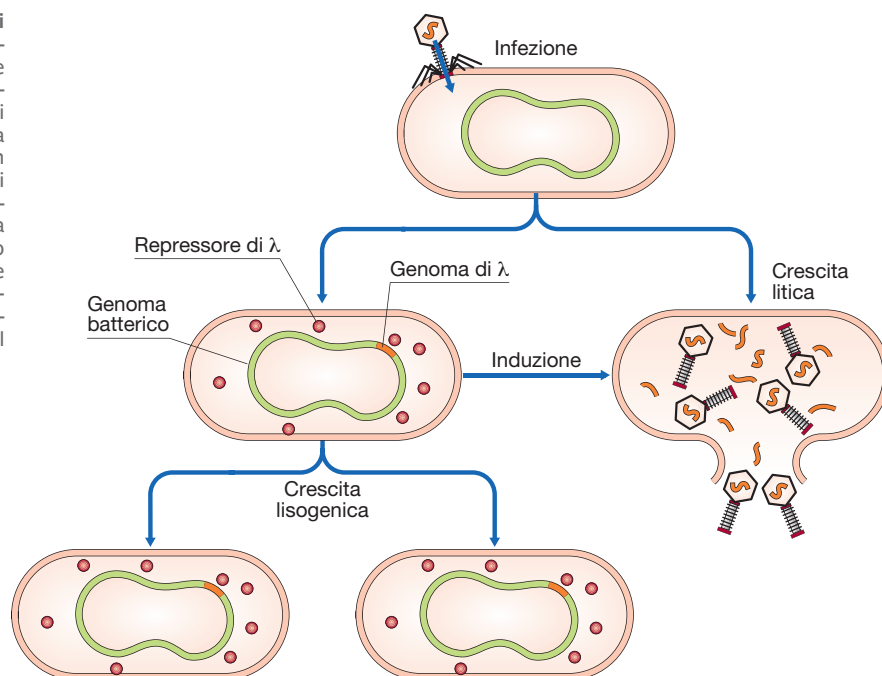


Figura 10.23 Mappa genomica del fago lambda nella sua forma circolarizzata. Al centro in alto è indicata la "regione di controllo", chiamata anche "regione di immunità", che contiene le

sequenze implicate nel controllo del ciclo litico e lisogenico del fago.

forma lineare ma, dopo l'entrata nel batterio ospite, la molecola lineare circola- rizza per interazione tra le estremità "coesive" costituite da due sequenze comple- mentari che protrudono alle estremità (*cos*). La regione centrale del cromosoma, detta regione di controllo o anche di immunità, contiene i geni che codificano per le proteine regolatrici dei due cicli, litico o lisogenico, di cui parleremo più avanti. Ai lati di questa si trovano due regioni contenenti rispettivamente alcuni geni che codificano gli enzimi implicati nella ricombinazione, integrazione ed escissione del DNA (a sinistra), e geni codificanti proteine replicative (a destra). Infine ci sono, raggruppati in quasi metà del genoma, circa 20 geni che codifi- cano per le proteine strutturali della testa e della coda del fago e due geni per gli enzimi necessari per la lisi batterica.

Ciclo litico

Come abbiamo detto, intraprendendo il ciclo litico il fago lambda, dopo avere infettato la cellula, utilizza l'apparato sintetico dell'ospite per promuovere la produzione di nuovi fagi, fino alla lisi della cellula batterica e il rilascio della progenie fagica. Nel corso di questo ciclo, che dura circa 20 minuti, vengono espressi in successione temporale, come succede per tutti i fagi litici, tre gruppi di geni, che nel caso di lambda vengono chiamati geni precoci immediati, geni precoci ritardati e geni tardivi. La transizione temporale della trascrizione da uno all'altro di questi gruppi genici, che in SPO1 e altri fagi è dovuta, come già detto, a una cascata di fattori σ alternativi, nel caso del fago lambda è dovuta ad antiterminazione, un meccanismo che opera a livello della terminazione della trascrizione.

La **figura 10.24** mostra una mappa parziale e schematizzata del cromoso- ma del fago lambda con, al centro, il gene *cI* che codifica per il "repressore di lambda" attivo durante il ciclo lisogenico. A sinistra e a destra di *cI* si trovano i due geni precoci immediati *N* e *cro*, che codificano per due regolatori della trascrizione che, come vedremo, operano rispettivamente sulla terminazione e sull'inizio. Questi sono orientati in modo divergente e perciò trascritti sui due diversi filamenti della doppia elica. Subito dopo *N* e *cro* ci sono due termina- tori della trascrizione, attivi durante i primi minuti del ciclo litico, che causano la terminazione della trascrizione e, quindi, l'espressione di solo questi due geni (tratto rosso dei due trascritti). Il prodotto del gene *N*, la proteina pN, è un antiterminatore che inibisce la terminazione della trascrizione alla fine di *N* stesso e di *cro*, permettendo così la trascrizione anche dei geni precoci ritardati, a valle in ambedue le direzioni (**fig. 10.24**). A valle di *N* si trovano il gene *cIII* e i geni responsabili della ricombinazione che permettono al cromosoma di

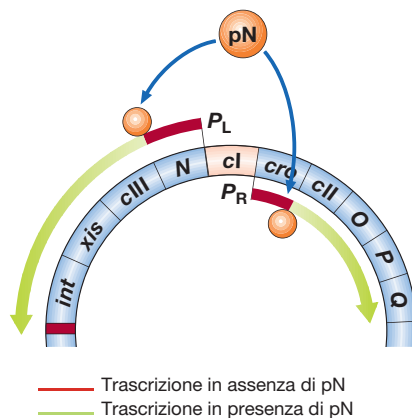


Figura 10.24 Eventi precoci che portano all'instaurazione del ciclo lisogenico. La proteina pN, codificata dal gene *N*, permette la prosecuzione della trascrizione, verso destra, oltre il gene di *Cro*, dei geni *cII*, *O*, *P* e *Q*, e, verso sini- stra, oltre il gene *N*, dei geni *cIII*, *xis* e *int*. P_R , promotore destro; P_L : promotore sinistro.

lambda di integrarsi o escindersi dal cromosoma di *E. coli*. A valle di *cro* troviamo il gene per *cII* e i geni *O*, *P* e *Q* implicati nella replicazione. Tra questi, il gene *Q*, che si trova verso la fine dell'operone di destra, produce una proteina, pQ, che a sua volta è un antiterminatore che permette l'ulteriore prosecuzione della trascrizione che porterà alla sintesi delle proteine tardive, le quali sono sostanzialmente le proteine strutturali della testa e della coda del fago; infine troviamo i geni per gli enzimi che porteranno alla lisi della cellula e al rilascio della progenie fagica.

Come funzionano pN e pQ? Per quanto riguarda pN, l'antiterminazione dipende dalla formazione di un complesso ribonucleoproteico sulla superficie dell'RNA polimerasi. Inizialmente, come mostrato in **figura 10.25** per l'operone di sinistra, l'RNA polimerasi trascrive solo il gene *N* fino al codone di terminazione alla fine di questo gene (**fig. 10.25A**). L'mRNA così prodotto dirigerà la sintesi della proteina pN, che, successivamente, insieme a fattori batterici del "complesso Nus" (come NusA, NusB, NusE e NusG) si lega agli mRNA nascenti per pN, in un sito particolare chiamato "nut" (*N utilization* = utilizzazione di N). Questo sito si trova 60 nt dopo l'inizio della trascrizione. Una volta legata all'mRNA, pN interagisce con l'RNA polimerasi impedendole di riconoscere il terminatore, che in questo caso è Rho-dipendente, e permettendole così di proseguire la trascrizione dei geni a valle (**fig. 10.25B**). La situazione è molto simile per l'operone di destra. Infatti, inizialmente viene trascritto solo il gene *cro*; con la sintesi di pN e il legame di questa su un sito nut, si ha l'inibizione della terminazione e la prosecuzione della trascrizione dei geni a valle.

L'altro antiterminatore, la proteina pQ, funziona in modo un po' diverso. Infatti essa, a differenza di pN, non riconosce una sequenza sull'RNA, ma una sequenza

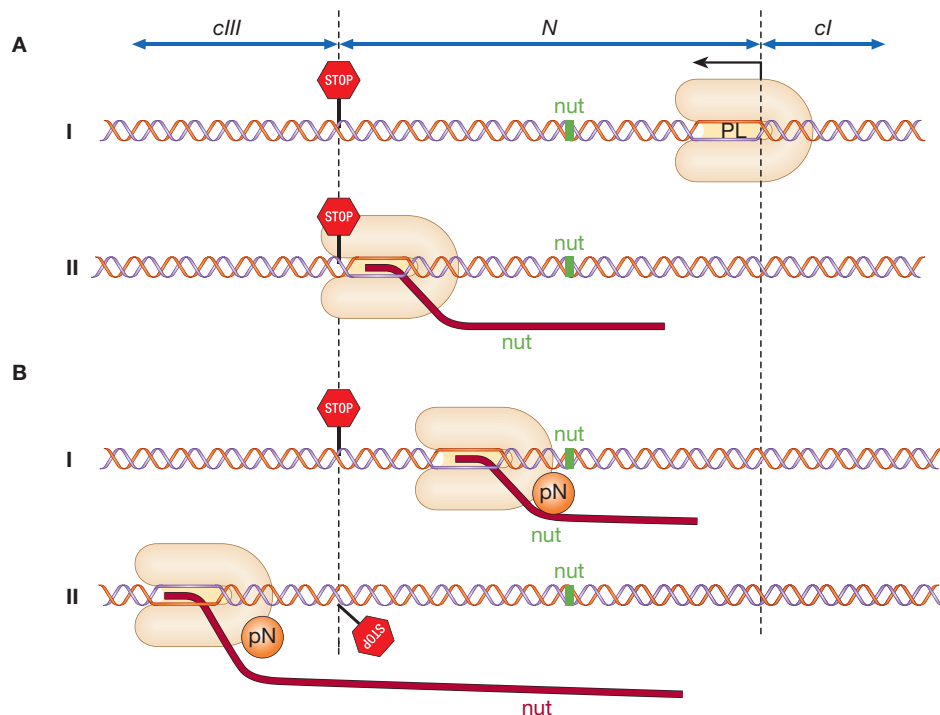


Figura 10.25 pN controlla la trascrizione dei geni precoci mediante un meccanismo di antiterminazione. (A) Dopo l'infezione, in assenza della proteina pN, l'RNA polimerasi inizia a trascrivere il gene *N* (AI), ma si ferma poi al terminatore della trascrizione (stop) posto a valle del gene *N* stesso (AII), portando alla produzione della proteina pN. (B) Raggiunta una concentrazione

sufficiente, pN si lega, come complesso comprendente proteine della famiglia Nus, a una sequenza dell'RNA nascente chiamata nut (indicata qui sul DNA come box verde) e interagisce poi con l'RNA polimerasi (BI). In presenza di pN, l'RNA polimerasi non riconosce il segnale di stop e continua la trascrizione dopo il gene *N*, fino a oltre il gene *int* (BII).

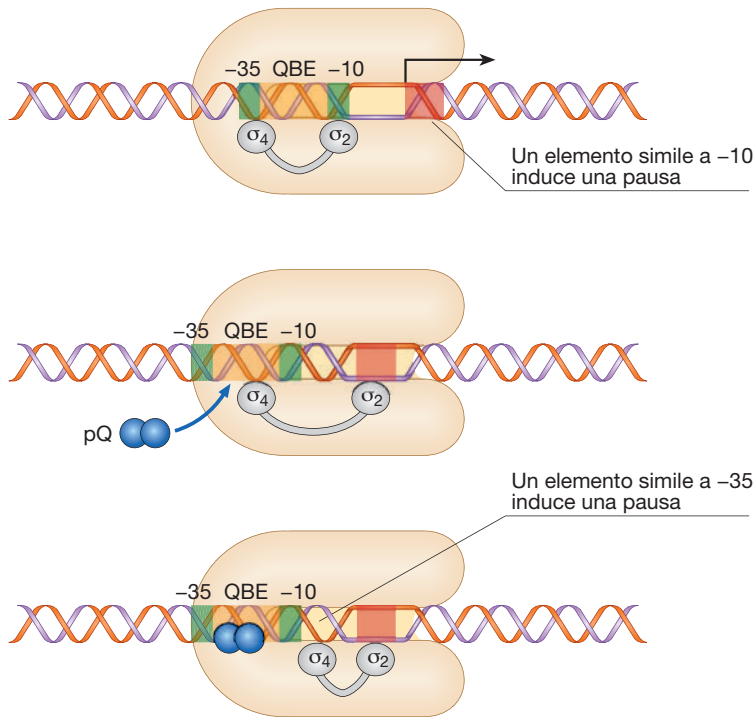


Figura 10.26 Meccanismo di antiterminazione mediata da pQ. La proteina pQ riconosce e lega le sequenze QBE (Q Binding Element) sul DNA, tra le regioni -10 e -35 del promotore dei geni tardivi. (A) In assenza di pQ la polimerasi inizia a trascrivere, ma si ferma dopo pochi nucleotidi perché σ_2 riconosce un elemento simile alla sequenza -10. Dopo la pausa la polimerasi si blocca al terminatore posto 200 pb a valle. (B) Se, durante la pausa, è presente pQ, questa si lega a QBE (C) e permette alla polimerasi di trascrivere oltre il terminatore. pQ interagisce probabilmente con la regione σ_4 modificando le interazioni con il nucleo della polimerasi. pQ apparentemente rimane legata alla polimerasi durante la trascrizione.

sul DNA chiamata QBE (Q Binding Element) e situata tra le regioni -10 e -35 del promotore dei geni tardivi. Come illustrato in **figura 10.26**, in assenza di pQ, la polimerasi inizia la trascrizione e si ferma al terminatore; in presenza di pQ, questo si lega alla sequenza QBE, interagisce con la polimerasi e la rende capace di trascrivere oltre il terminatore. Il meccanismo preciso con il quale funziona pQ non è noto, ma i modelli proposti suggeriscono che pQ interagisca con il fattore σ , e in particolare con σ_4 , nelle vicinanze del canale di uscita dell'RNA.

Ciclo lisogenico

Alternativamente, il fago lambda può seguire un'altra via, il ciclo lisogenico. Durante la crescita lisogenica il genoma di lambda viene integrato nel cromosoma della cellula ospite, replicato passivamente e propagato come un tratto del DNA del batterio. Alcuni aspetti del ciclo vitale di lambda e il meccanismo con il quale il suo DNA viene integrato nel cromosoma di *E. coli* per ricombinazione sito-specifica sono stati trattati nel paragrafo 8.2. Quando il genoma del fago è integrato nel cromosoma batterico è definito **profago** e il **batterio** che lo contiene si dice **lisogenico**; questa condizione è molto stabile e può andare avanti in questo stato per molte generazioni. L'esposizione del batterio lisogenico a radiazioni o ad altri agenti che danneggiano il DNA e che, quindi, influenzano la crescita delle cellule batteriche può attivare l'**induzione lisogenica**, che risulta nell'escissione, dal cromosoma di *E. coli*, del genoma fagico, che rientra nel ciclo litico.

Regione di controllo del fago

Vediamo ora come avviene la scelta tra i due cicli, analizzando gli elementi che controllano l'espressione dei geni di lambda. La regione di controllo, detta anche **regione di immunità**, illustrata in **figura 10.27**, è costituita da due geni, *cI* e *cro* (Control of Repressor and Other things), e tre promotori, P_R , P_L e P_{RM} .

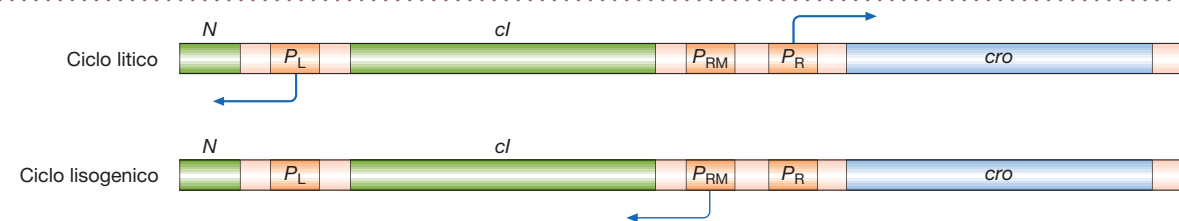


Figura 10.27 Regione di controllo e trascrizione nel ciclo litico e lisogenico. La “regione di controllo”, detta anche “regione di immunità”, è costituita da due geni, *cl* e *cro*, e tre promotori, P_R , P_L e P_{RM} . Le frecce mostrano quali promotori sono attivi per dirigere

lo sviluppo di lambda verso il ciclo litico, in alto, o lisogenico, in basso. Si noti che, nel ciclo lisogenico, l'unico promotore acceso è P_{RM} , che mantiene la sintesi del repressore.

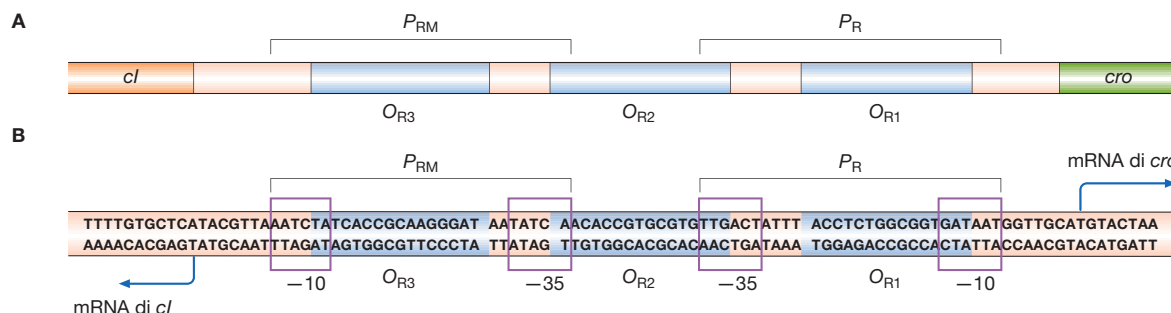


Figura 10.28 Il tratto centrale della regione di controllo. Il tratto di circa 80 pb compreso tra i geni *cl* e *cro* contiene i promotori divergenti P_{RM} e P_R . Sovrapposti con questi si trovano i tre siti

operatori O_{R1} , O_{R2} e O_{R3} . Le frecce blu indicano i siti d'inizio e la direzione della trascrizione.

I promotori P_R e P_L (destra e sinistra) sono responsabili della trascrizione dei geni precoci immediati *N* e *cro* che vengono trascritti subito dopo l'infezione e, indirettamente, sono responsabili della trascrizione di tutti gli altri geni di lambda. Si tratta di due promotori forti e costitutivi che legano l'RNA polimerasi dell'ospite senza bisogno di altri fattori di transattivazione. Il promotore P_{RM} , responsabile del mantenimento del repressore codificato dal gene *cl*, è, invece, un promotore debole e ha bisogno di un attivatore per reclutare la polimerasi. Come si vede nella figura, nel ciclo litico sono attivi i due promotori P_L e P_R , mentre nel ciclo lisogenico è attivo solo il promotore P_{RM} . Come vedremo, l'attivazione del promotore P_{RM} esclude la funzione di ogni altro promotore del fago lambda.

Guardiamo ora in dettaglio (fig. 10.28) come sono fatte le regioni di controllo e la sequenza che contiene, sovrapposti, i promotori divergenti P_R e P_{RM} e tre siti operatori, chiamati O_{R1} , O_{R2} e O_{R3} , lunghi ciascuno 17 pb. Questi vengono riconosciuti e legati dalle proteine regolatrici della trascrizione, il repressore di lambda (CI) e Cro. Nella parte inferiore della figura sono evidenziate le regioni -10 e -35 dei due promotori divergenti. È importante notare che l'operatore O_{R2} si sovrappone alle regioni -35 di entrambi i promotori e questo, come vedremo, permette al repressore di controllare la trascrizione di tutti e due i geni. Questi operatori di 17 pb hanno una diversa affinità per le proteine regolatrici, repressore e Cro, che a basse concentrazioni andranno a legarsi solo ai siti forti, mentre col crescere della loro concentrazione si legheranno anche ai siti più deboli.

Struttura del repressore e di Cro

Il gene *cl* codifica per il repressore del ciclo litico chiamato anche **repressore di lambda**. Come illustrato in figura 10.29, questa proteina è costituita da due domini che sono collegati tra loro da una regione flessibile (*linker*). La parte N-terminale

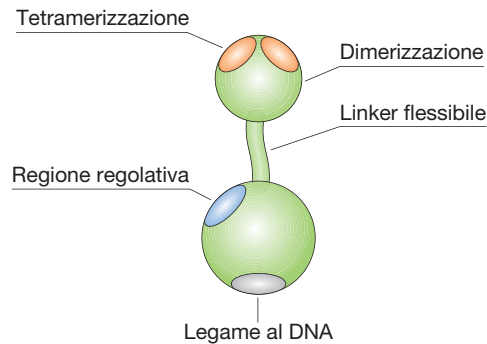


Figura 10.29 Rappresentazione schematica del repressore del fago lambda. Il repressore (CI) del fago lambda è costituito da due domini strutturali. Il *dominio C-terminale* contiene due regioni impegnate, rispettivamente, nella dimerizzazione e nella tetramerizzazione; il *dominio N-terminale* contiene il motivo strutturale elica-giro-elica per il legame al DNA e la regione attivatrice che interagisce con il fattore σ dell'RNA polimerasi. I due domini sono connessi da un "linker flessibile".

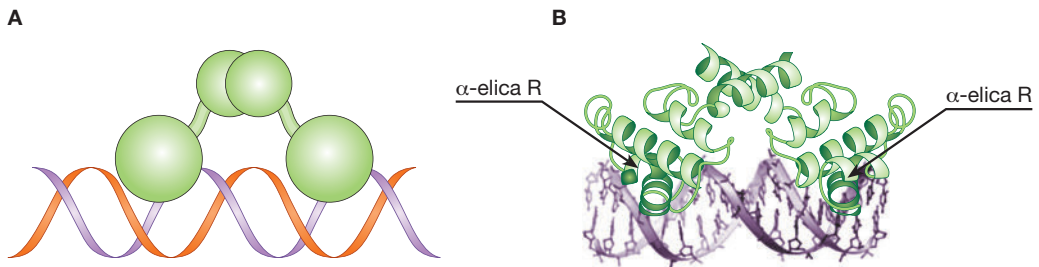


Figura 10.30 Interazione tra repressore di lambda e sito operatore. Il repressore CI si lega come dimero al sito operatore, mediante l'interazione specifica, nel solco maggiore della doppia elica

del DNA, tra l' α -elica R del motivo elica-giro-elica e la sequenza di controllo O, di 17 pb.

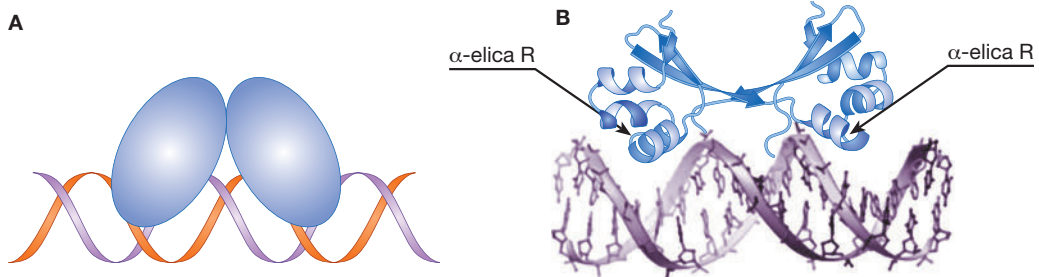


Figura 10.31 Cro, come il repressore di lambda, si lega al DNA in forma di dimero. L'interazione avviene nel solco maggiore della

doppia elica di DNA, tra l' α -elica R del motivo elica-giro-elica e la sequenza di controllo O, di 17 pb.

contiene il motivo strutturale elica-giro-elica che le permette di legarsi in modo specifico al DNA e una regione che viene definita attivatrice perché interagisce con la regione 4 della subunità sigma dell'RNA polimerasi, contribuendo al suo reclutamento sul promotore. Infatti, come abbiamo già detto, P_{RM} è un promotore debole che ha bisogno di attivazione per funzionare con efficienza. Il dominio C-terminale del repressore contiene una regione che gli permette di dimerizzare; infatti il repressore di lambda si lega al DNA come dimero (**fig. 10.30**); in particolare si lega con i motivi elica-giro-elica ai siti di 17 pb O_{R1} , O_{R2} e O_{R3} . Come si può vedere, l' α -elica R (di riconoscimento) interagisce in modo specifico con il solco maggiore del DNA. Nel dominio C-terminale è presente anche un'altra regione che permette al repressore di tetramerizzare e, come vedremo successivamente, questa caratteristica gli consente di legarsi in modo cooperativo al DNA. Un aspetto interessante della funzione del repressore è che essa è duplice: da una parte, reprime tutti i geni di lambda con un meccanismo di controllo negativo; dall'altra, attiva la sua stessa espressione con un meccanismo di controllo positivo.

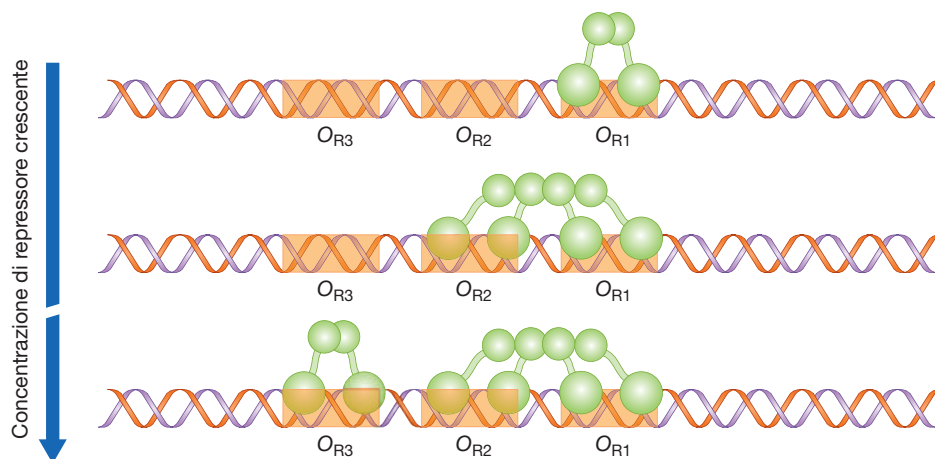


Figura 10.32 Il repressore si lega ai tre siti operatori con diversa affinità. A bassa concentrazione, il repressore si lega come dimeri al sito ad alta affinità O_{R1} ; a concentrazioni intermedie si ha il legame di un secondo dimeri al sito O_{R2} ; questo secondo legame

è di per sé a bassa affinità, mentre è alta l'affinità tra il tetramero e i due siti O_{R1} e O_{R2} . A concentrazioni di repressore ancora maggiori un terzo dimeri si lega al sito O_{R3} .

L'altra proteina che agisce su questa regione di controllo è Cro (**fig. 10.31**), che ha soltanto la funzione di repressore; si lega al DNA come un dimeri e impedisce il legame del repressore nella regione di controllo (**fig. 10.32**). Anche Cro lega il DNA con un motivo elica-giro-elica; in particolare l' α -elica R riconosce specificamente il solco maggiore del DNA nei tre operatori.

CI e Cro si legano con diversa affinità alla regione di controllo

I siti di 17 pb O_{R1} , O_{R2} e O_{R3} hanno sequenze simili ma non identiche tra loro e questa caratteristica permette il legame al repressore e a Cro con affinità molto diverse. Il repressore ha un'affinità per O_{R1} dieci volte maggiore che per gli altri due siti e, come si può vedere in **figura 10.32**, a bassa concentrazione il repressore, in forma dimerica, sarà legato solo a O_{R1} . A concentrazioni leggermente più elevate il repressore si legherà anche a O_{R2} sfruttando la sua capacità di formare tetrameri, grazie alla flessibilità della regione linker. In altre parole, il legame a O_{R2} è cooperativo, nel senso che a queste concentrazioni intermedie il repressore non sarebbe in grado di legare O_{R2} se O_{R1} non fosse già legato a una molecola di repressore. A concentrazioni molto elevate il repressore si legherà anche a O_{R3} .

Cro ha, invece, un'affinità 10 volte maggiore per O_{R3} che per gli altri due siti ma, data la sua struttura, non è in grado di formare legami cooperativi per aiutare un altro dimeri di Cro a legare il sito adiacente. La **figura 10.33** illustra il legame di Cro alla regione di controllo al crescere della sua concentrazione. Quindi Cro e il repressore hanno un'affinità opposta per i siti operatori della regione di controllo, con il repressore che ha come sito ad alta affinità O_{R1} e Cro che ha come sito ad alta affinità O_{R3} . La **figura 10.34** mostra le basi strutturali della differenza di legame del repressore e di Cro ai loro rispettivi siti ad alta affinità. Come si può vedere, sia nel repressore che in Cro le α -eliche di riconoscimento iniziano entrambe con gli amminoacidi glutammina-serina che interagiscono con alcune paia di basi comuni nella sequenza di riconoscimento, come A-T in posizione 2 e C-G in posizione 4. Queste basi sono altamente conservate in tutti i siti di riconoscimento della regione di controllo e questo permette alle due proteine di legarsi a tutti i siti operatori. Le differenti affinità sono invece dovute al resto dell' α -elica e alla sequenza di riconoscimento, dove si possono avere ulteriori contatti che aumentano la specificità e la forza dell'interazione DNA-proteina.

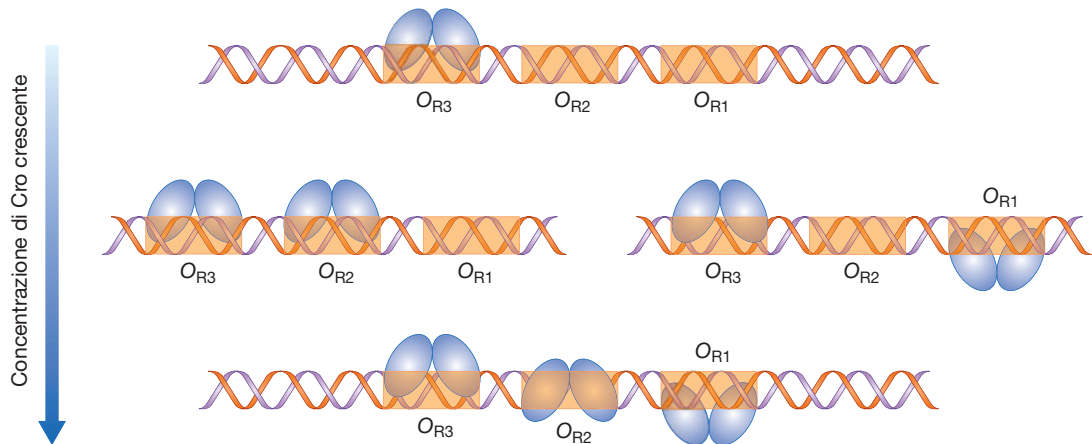


Figura 10.33 Cro si lega ai tre siti operatoriali con diversa affinità. Al contrario del repressore, Cro presenta una maggiore affinità per O_{R3} che per gli altri siti O_{R2} e O_{R1} . A concentrazioni un po' più elevate un secondo dimerio di Cro si legherà indifferentemente a O_{R2}

e O_{R1} ; infatti, non essendo in grado di formare tetrameri, Cro non può legarsi in modo cooperativo a due siti O_R . A concentrazioni ancora più elevate Cro si legherà su tutti e tre i siti O_R .

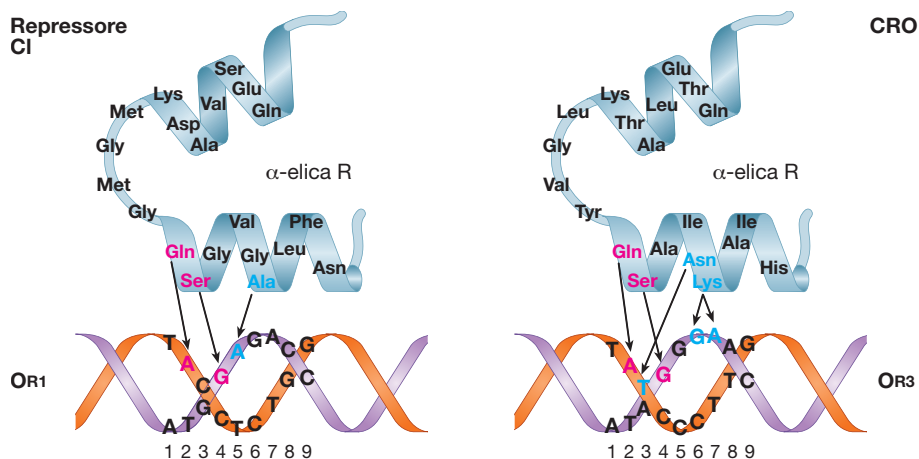


Figura 10.34 Rappresentazione strutturale semplificata delle interazioni del repressore e di Cro con i rispettivi siti di maggiore affinità. Le sequenze O_{R1} e O_{R3} presentano alcune basi conservate e altre diverse. Analogamente, le α -eliche di riconoscimento dei motivi elica-giro-elica del repressore e di Cro hanno alcuni residui identici e

altri diversi. Questo permette alle due proteine di legarsi comunque ai due siti utilizzando le caratteristiche comuni, ma di mantenere una diversa affinità per i due operatori. In rosso e in blu sono indicati, rispettivamente, amminocidi e basi conservati e differenti nelle due interazioni. Le frecce indicano i contatti fra le proteine e il DNA.

Il repressore CI e Cro controllano in modo mutuamente esclusivo la trascrizione nella regione di controllo del fago lambda

Vediamo ora quali sono gli effetti, sull'espressione dei geni di lambda, delle differenti affinità nel riconoscimento dei siti di controllo da parte del repressore e di Cro. Osserviamo attentamente, in **figura 10.28**, la struttura della regione di controllo e, in particolare, dei due promotori divergenti, P_R e P_{RM} dei geni *cro* e *cI*. Nel caso di *cI* sappiamo che il promotore è debole e ha bisogno di un attivatore per essere trascritto. Il repressore di lambda, come abbiamo detto precedentemente, contiene nel dominio N-terminale una regione che interagisce con il fattore σ della polimerasi e la recluta sul promotore. Il repressore è, quindi, l'attivatore della sua stessa sintesi. Questo avviene quando il repressore è legato a O_{R1} e O_{R2} ed è la molecola di repressore in O_{R2} che ha la funzione attivatrice reclutando la polimerasi. Il legame cooperativo aiuta ad avere i due siti O_{R1} e O_{R2} occupati a livelli relativamente bassi di repressore. Se tuttavia i livelli di repressore

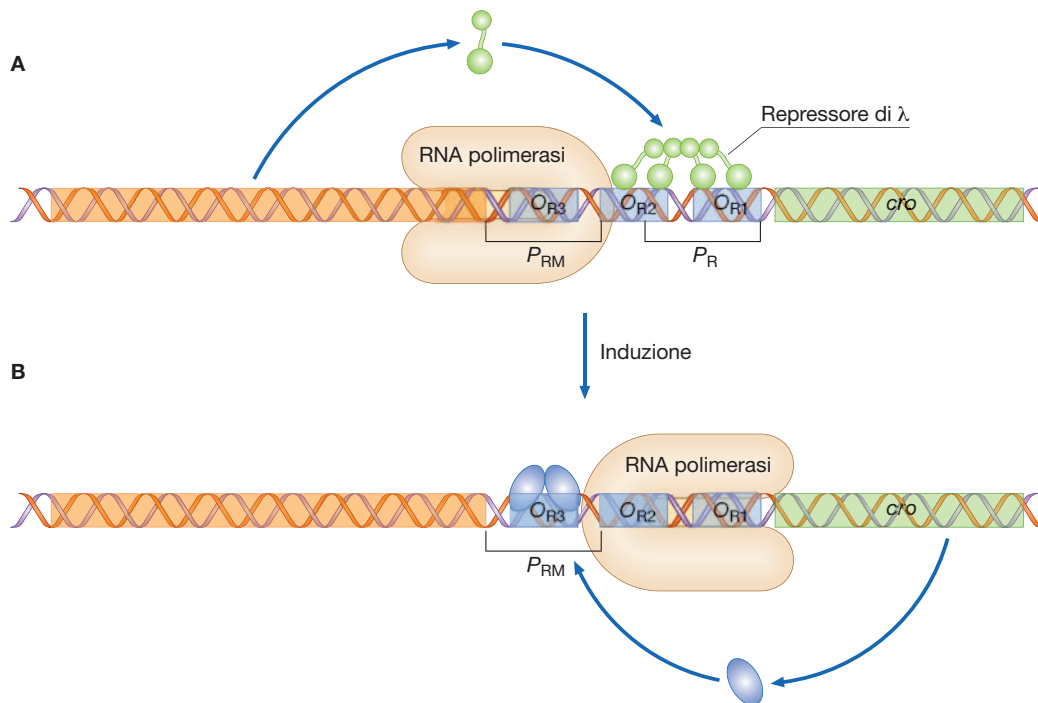


Figura 10.35 Azione del repressore di lambda e di Cro. (A) Quando il repressore CI è legato a O_{R1} reprime la trascrizione del gene *cro* a partire dal promotore P_R ; se è legato anche a O_{R2} , continua a reprimere *cro* ma attiva la trascrizione di se stesso dal promotore P_{RM} . Questa condizione è caratteristica del ciclo lisogenico quando tutti gli altri geni di lambda sono spenti. A concentrazioni ancora più

elevate il repressore si lega anche al sito O_{R3} , inibendo così anche la trascrizione di se stesso. (B) Nel ciclo litico Cro prende il sopravvento e legandosi al sito O_{R3} , reprime la sintesi del repressore CI a partire dal promotore P_{RM} . Quando Cro raggiunge concentrazioni più elevate si lega anche ai siti O_{R2} e O_{R3} , inibendo anche la trascrizione di se stesso.

si alzano troppo, interviene un'autoregolazione negativa, perché anche il sito O_{R3} sarà occupato dal repressore CI e questo reprimerà il promotore P_{RM} . Questo meccanismo regola in modo molto fine i livelli di repressore presenti nella cellula durante la lisogenia. Nello stesso tempo, il repressore, legandosi ai siti O_{R1} e O_{R2} , impedisce il legame della polimerasi sul promotore P_R , spegnendo la trascrizione di Cro e di tutti i geni a valle regolati da questo promotore.

Anche Cro regola la trascrizione della regione di controllo, ma lo fa usando solo un controllo negativo. Quando lambda infetta le cellule di *E. coli*, il promotore P_R di Cro è libero e la polimerasi trascrive il messaggero di Cro. La proteina prodotta va a legarsi al sito con più alta affinità, che è O_{R3} , impedendo alla polimerasi di trascrivere dal promotore P_{RM} e inibendo così la produzione del repressore. Se i livelli di Cro sono troppo elevati, questa molecola si legherà anche a O_{R1} , spegnendo il promotore P_R , che è responsabile della sua trascrizione, anche in questo caso usando un sistema di autoregolazione negativa. La **figura 10.35** sintetizza quanto abbiamo appena detto. In alto si vede la regione di controllo nella fase lisogenica, con il repressore legato a O_{R1} e O_{R2} che attiva la sua stessa sintesi, impedendo a Cro di legarsi a O_{R3} e di inibire il promotore P_{RM} . In basso, invece, è descritto il controllo durante il ciclo litico, dove Cro viene prodotto e, legandosi a O_{R3} , blocca la trascrizione del gene *cI* inibendo la produzione di repressore.

Legame cooperativo del repressore al DNA

Abbiamo detto che il repressore si lega ai siti di regolazione O_{R1} e O_{R2} in modo cooperativo e che è necessaria una quantità molto maggiore di proteina per legarsi a O_{R3} e inibire così la propria trascrizione. In realtà lambda usa un sistema ancora più

raffinato che gli permette di dosare in modo molto accurato la quantità di repressore presente nella cellula. A valle del gene *cI* esistono altri tre siti operatori, O_{L1} , O_{L2} , e O_{L3} , parzialmente sovrapposti con il promotore di sinistra P_L che controlla la trascrizione del gene *N*. Gli operatori O_{L1} , O_{L2} e O_{L3} sono uguali come sequenza e invertiti come orientamento rispetto a O_{R1} , O_{R2} e O_{R3} , che permettono il legame del repressore. Come illustrato in **figura 10.36**, le molecole di repressore legate a O_{R1} e O_{R2} interagiscono con i dimeri legati su O_{L1} e O_{L2} . Questo legame è reso possibile dalla formazione di un'ansa di DNA nella regione tra questi operatori, ansa stabilizzata dal legame dei repressori legati ai propri siti forti e che formano un ottamero. Questa situazione avvicina tra loro anche i siti deboli O_{L3} e O_{R3} , che possono legare in modo cooperativo altri due dimeri del repressore. Grazie a questo nuovo legame cooperativo il sito O_{R3} , necessario per l'autoregolazione negativa, viene occupato a una concentrazione di repressore significativamente più bassa di quella che sarebbe necessaria in assenza di cooperatività. La conseguenza di questo legame rende la concentrazione intracellulare di repressore circa tre volte più bassa e questo è molto importante per l'induzione del ciclo litico in un batterio lisogenico per lambda. Il grafico di **figura 10.37** mostra come la curva dell'efficienza di legame del repressore al DNA ha un andamento sigmoide tipico dei legami cooperativi. La linea tratteggiata rappresenta il legame del repressore su un singolo sito, la curva in rosso il legame cooperativo a due siti. Come si vede, la percentuale di DNA legato aumenta in modo più rapido.

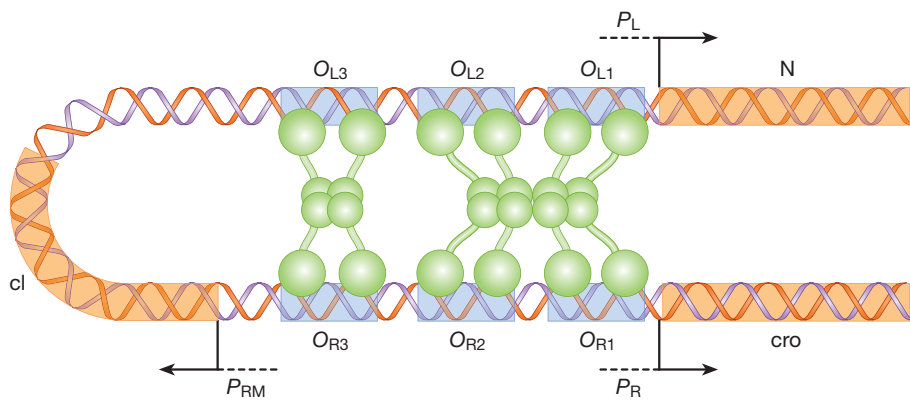


Figura 10.36 Interazione a distanza degli operatori O_L e O_R . A valle del gene *cI* (cioè a sinistra di *cI* nella nostra rappresentazione del cromosoma di lambda) è presente un altro operatore chiamato O_L , anche questo costituito da 3 siti operatori, O_{L1} , O_{L2} e O_{L3} , con la stessa sequenza nucleotidica di O_R , ma invertita. I dimeri di repressore legati a O_{R1} e O_{R2} interagiscono con i dimeri legati a O_{L1} e O_{L2} , risultando in un legame cooperativo ad alta affinità tra

l'ottamero di repressore e i due siti operatori di destra e di sinistra. Questa interazione comporta il ripiegamento ad ansa della doppia elica di DNA che permette l'avvicinamento degli operatori O_R e O_L e il legame di altri due repressori dimerici ai siti di O_{R3} e O_{L3} . Questi siti di legame, singolarmente deboli, incrementano però ulteriormente l'affinità del repressore multimerico ai siti operatori.

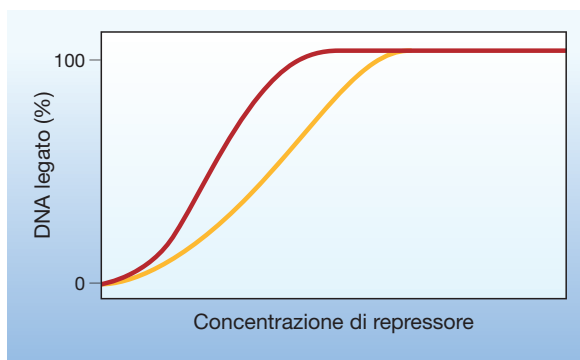


Figura 10.37 Legame cooperativo del repressore di lambda. La linea arancione e la linea rossa descrivono l'andamento del legame del repressore rispettivamente a un singolo sito e a due siti operatori. Si può notare come, nel secondo caso, il legame al DNA è cooperativo, cioè molto più rapido e viene raggiunto a concentrazioni di repressore notevolmente più basse.

Induzione del ciclo litico di lambda in un batterio lisogenico

Come abbiamo detto precedentemente, se un batterio che contiene nel suo cromosoma un genoma di lambda integrato come profago e mantenuto stabilmente come lisogenico viene irradiato con raggi ultravioletti, si ha induzione della crescita del fago, cioè escissione del genoma di lambda dal cromosoma batterico, produzione di nuove particelle e lisi della cellula. Questo fenomeno è scatenato dalla degradazione del repressore da parte del sistema di proteine che rispondono a danni al DNA, in quella che viene chiamata la **risposta SOS** (vedi anche finestra 7.1). Infatti, nel caso di danni al DNA del batterio è interesse del fago sfuggire a una situazione potenzialmente pericolosa. Nella risposta SOS si attiva una proteina batterica denominata **RecA**, che normalmente ha un ruolo nella ricombinazione. In questo caso RecA ha la funzione di permettere l'autodegradazione di un repressore batterico che si chiama **LexA**, che reprime la sintesi dei geni responsabili della riparazione del DNA. Il repressore di lambda si è evoluto per assomigliare a LexA e quindi è un bersaglio dell'attività proteolitica di RecA, come è illustrato in **figura 10.38**. Il taglio proteolitico sul repressore separa il dominio di legame al DNA dal dominio C-terminale e, quindi, abolisce la dimerizzazione e il legame cooperativo. Ciò ha come effetto il distacco del repressore dai siti operatori (**fig. 10.39**) liberando quindi il promotore P_R . A questo punto può riprendere la sintesi di Cro, che reprimerà a sua volta la sintesi del repressore CI. È importante

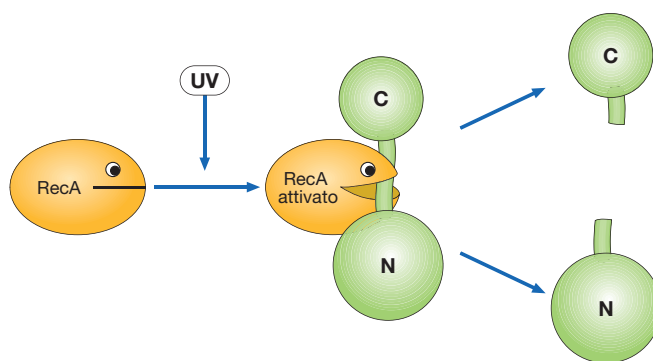
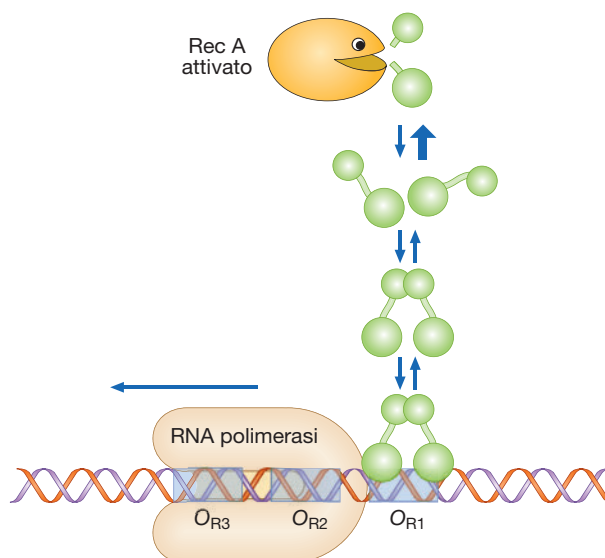


Figura 10.38 Induzione del ciclo litico di lambda in un batterio lisogenico. I raggi ultravioletti (UV) attivano la proteina RecA che

causa la proteolisi del repressore tagliando il linker flessibile che risulta nella separazione dei due domini N-terminale e C-terminale.

Figura 10.39 La degradazione del repressore promuove il suo distacco dall'operatore. La separazione dei domini C- e N-terminale del repressore CI ha come effetto la perdita dei domini di dimerizzazione del repressore stesso, che non è più in grado di legarsi cooperativamente al DNA. Le coppie di frecce blu indicano lo stabilirsi di equilibri dinamici tra le varie forme di repressore; la freccetta più spessa (in alto a destra) indica lo spostamento dell'equilibrio dovuto al taglio irreversibile della regione linker del repressore.



che i livelli di repressore siano regolati in modo fine per far funzionare efficientemente l'induzione del ciclo litico. È per questo che lambda ha evoluto il sistema dei legami cooperativi: per ridurre i livelli di repressore necessari a mantenere lo stato lisogenico e permettere un'induzione veloce per fuggire da un batterio con il DNA danneggiato.

Ruolo dell'attivatore trascrizionale CII

Come descritto nella **finestra 10.3**, l'isolamento e lo studio di mutanti che, non essendo in grado di lisogenizzare, intraprendono esclusivamente il ciclo litico formando placche chiare su una piastra di agar, hanno portato all'isolamento dei geni *cI*, *cII* e *cIII* (*c* sta per *clear*, facendo riferimento alla placca chiara). Sappiamo che *cI* codifica il repressore di lambda e, quindi, i mutanti che ne limitano la funzione hanno come conseguenza la crescita litica. I geni *cII* e *cIII* codificano per due proteine che hanno funzioni diverse: CII è un attivatore della trascrizione per promotori deboli che hanno bisogno di un controllo positivo e CIII è una proteina che protegge CII dalla degradazione proteolitica. Vediamo ora come sono organizzati questi geni sul cromosoma di lambda. Come si può vedere in **figura 10.40**, *cII* si trova a valle del gene *cro* ed è trascritto insieme a questo verso destra a partire dal promotore P_R , mentre *cIII* è a valle del gene *N* ed è trascritto verso sinistra a partire dal promotore P_L . Come abbiamo detto, CII è un attivatore della trascrizione e agisce su vari promotori deboli di lambda che sono sotto il suo controllo positivo.

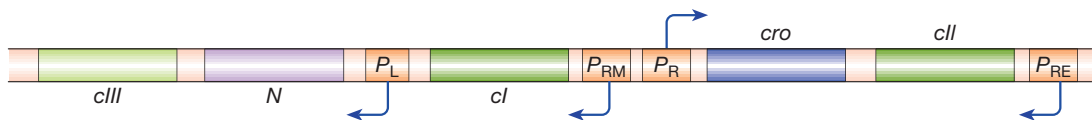


Figura 10.40 Regione del cromosoma di lambda contenente i promotori che controllano i cicli del fago. Questo tratto genomico contiene la regione di controllo e i geni fiancheggiati, *N* e *cIII* a sinistra e *cI* a destra. Inoltre sono indicati i quattro promotori: P_L per la trascrizione di

N e, per effetto dell'antiterminazione, anche dei geni a sinistra, *cIII* e seguenti; P_R per la trascrizione di *cro* e, per effetto dell'antiterminazione, anche dei geni a destra, *cII* e seguenti; P_{RE} per dare inizio alla trascrizione di *cI*; P_{RM} per il mantenimento della trascrizione di *cI*.

FINESTRA 10.3

IDENTIFICAZIONE DI MUTAZIONI NELLA REGIONE DI CONTROLLO

Gli studi che hanno portato alla definizione delle regioni di controllo del fago sono stati fatti negli anni sessanta del secolo scorso, con raffinati esperimenti di Genetica, individuando mutazioni che alteravano o stimolavano l'uno o l'altro dei due cicli vitali del fago. La **figura 10F3.1** mostra una piastra di agar dove sono stati seminati dei batteri infettati da lambda. Normalmente i batteri crescono sulla piastra formando un tappeto confluyente sulla sua superficie. Dove i fagi hanno infettato i batteri si vedono, sullo sfondo nero, due tipi di placche, alcune opache e altre limpide. Le placche opache sono dovute al fatto che alcuni fagi intraprendendo il ciclo lisogenico non lisano i batteri, che continuano a crescere. Se si prelevano dei batteri da una placca opaca e si riseminano su un'altra piastra, si osserva che questi batteri crescono normalmente e sono immuni da altre infezioni da parte di lambda. Questi batteri contengono il profago e se vengono irradiati produrranno una nuova progenie di fagi che liseranno il batterio. Le placche limpide, invece, sono dovute a una forma mutata di lambda che non è in grado di lisogenizzare, cioè integrare il proprio genoma in quello della cellula ospite; poiché questi fagi mutanti sono in

grado di portare avanti esclusivamente il ciclo litico, non ci saranno batteri lisogeni contenenti il profago e, di conseguenza, le placche saranno limpide.

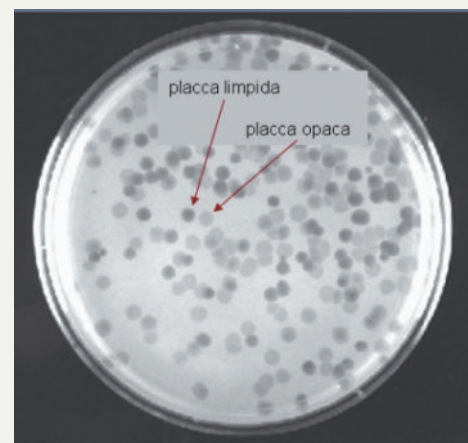


Figura 10F3.1 Placche di lisi fagiche.

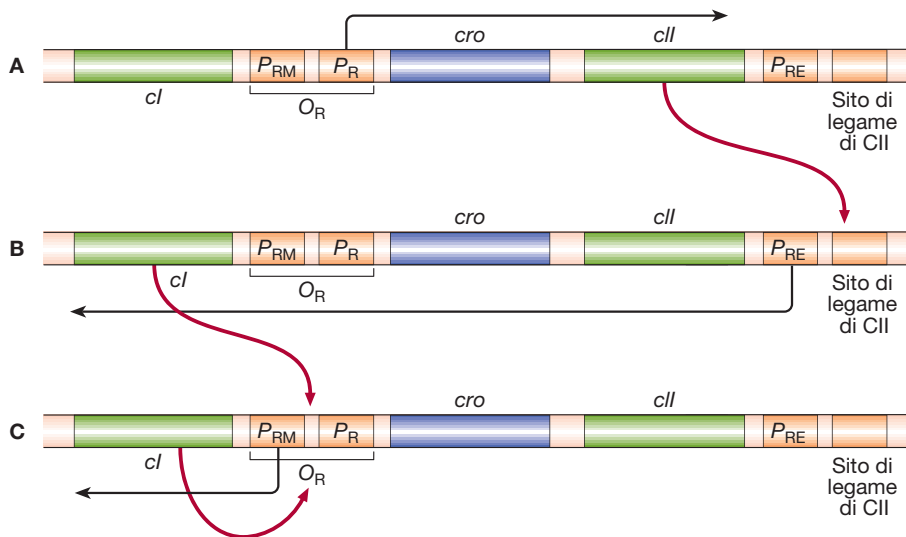


Figura 10.41 Processo d'instaurazione della lisogenia. (A) La trascrizione precoce di Cro inizia dal promotore P_R e, grazie al processo di antiterminazione mediato da pN, prosegue sui geni a valle, primo dei quali è cII ; (B) CII è un fattore di trascrizione positivo che, legandosi al suo sito bersaglio vicino al promotore P_{RE} , dà ini-

zio alla trascrizione di un RNA antisenso dei geni cII e cro , proseguendo come RNA senso di cI ; (C) il repressore cI così prodotto si lega al sito operatore O_R sovrapposto al promotore P_{RM} , attiva la trascrizione di mantenimento di cI e blocca quella di Cro.

Uno di questi è il promotore P_{RE} che si trova a valle di cII ma è orientato verso sinistra: esso promuove la sintesi delle prime molecole di repressore che vengono prodotte dopo l'infezione. Infatti, la funzione di P_{RM} richiede la presenza del repressore e quindi non può innescare inizialmente la trascrizione di cI , ma può "mantenere" una sintesi già iniziata da P_{RE} (P_{RE} sta per *Promoter for Repressor Establishment*: promotore per stabilire la sintesi del repressore). La **figura 10.41** mostra come la trascrizione, che parte dal promotore P_R , porta alla produzione di Cro e CII; quest'ultima proteina va ad attivare la trascrizione del repressore legandosi a un sito specifico a monte di P_{RE} . Infatti, CII interagisce direttamente con l'RNA polimerasi e la recluta sul promotore P_{RE} . La trascrizione da questo promotore usa come stampo il filamento opposto a quello usato per cro ; una volta che sarà stato sintetizzato un numero sufficiente di molecole di repressore, questo sarà in grado di legarsi agli operatori O_R e O_L e di attivare la propria sintesi dal promotore P_{RM} . Quindi, la sintesi del repressore è iniziata da P_{RE} ed è mantenuta da P_{RM} .

Cerchiamo ora di capire i meccanismi che inducono il fago lambda a scegliere, dopo l'infezione della cellula ospite, se proseguire con il ciclo litico o con il ciclo lisogenico. Abbiamo già detto che, subito dopo l'infezione di *E. coli*, il cromosoma lineare di lambda circolarizza grazie alle sue estremità, che, terminando con tratti a singolo filamento complementari, possono legarsi tra loro (sito *cos*). Abbiamo anche detto che subito inizia la trascrizione dei geni precoci immediati *N* e *cro* a partire dai promotori P_L e P_R . Questa prima attivazione genica dà il via al ciclo litico: antiterminazione da parte di pN, trascrizione dei geni precoci ritardati e seconda antiterminazione da parte di pQ con trascrizione dei geni tardivi. Se però durante la fase tardiva del ciclo litico viene attivata la trascrizione del gene *cI* a partire dal promotore P_{RE} , può essere attivato lo stato lisogenico. Questo promotore ha una sequenza -10 molto diversa dalla sequenza consenso canonica e, inoltre, manca della sequenza a -35. Questa struttura non ottimale del promotore rende indispensabile il ruolo di CII come transattivatore positivo. Inoltre, l'efficienza di traduzione dell'mRNA per *cI*, trascritto a partire da P_{RE} , è molto più alta di quella dell'mRNA per *cI* trascritto dal promotore P_{RM} . Infatti, il trascritto da P_{RE} ha un sito di legame al ribosoma molto forte, mentre il trascritto dal promotore P_{RM} ne

è privo e inizia la traduzione alla prima tripletta AUG. Questa situazione riflette la necessità di sintetizzare subito, all'inizio dell'infezione, numerose molecole di repressore per competere con la proteina Cro per i siti di controllo e nello stesso tempo di esprimere il repressore con moderazione, nella fase iniziale di crescita lisogenica, per permettere eventualmente una rapida ripresa del ciclo litico.

Da quanto abbiamo detto, è chiaro che la proteina CII è un fattore chiave per decidere quale dei due cicli verrà intrapreso; CII è il vero interruttore che dirige lambda verso il ciclo litico o il ciclo lisogenico. Se la proteina CII è presente in quantità elevate, il fago infettante andrà verso la lisogenia, mentre se non sarà abbondante e in grado di indurre la produzione di buone quantità di repressore, il fago andrà verso il ciclo litico. Le condizioni di crescita del batterio sono quelle che decidono quale delle due vie di sviluppo intraprendere. Infatti, quando i batteri sono sani e crescono in maniera efficiente, lambda tende a intraprendere la via litica, dove le nuove particelle fagiche possono infettare altre cellule sane di *E. coli*. Quando invece le cellule sono deboli e crescono lentamente, il fago tende a integrarsi e aspettare tempi migliori. La proteina CII, che è l'interruttore delle due vie di sviluppo, è molto instabile e viene degradata da una proteasi specifica che si chiama FtsH, che è codificata dal gene *hfl* di *E. coli*. La capacità di CII di indurre il ciclo lisogenico dipende dalla velocità con la quale viene degradata dalla proteasi. Infatti, se le cellule crescono bene e con efficienza, le proteasi sono più abbondanti, CII scarseggerà e il fago proseguirà nel ciclo litico. Se, al contrario, le cellule sono deboli e poco attive nel loro metabolismo, le proteasi saranno meno abbondanti e questo darà il tempo a CII di attivare la sintesi del repressore e la via lisogenica. Il gene *cIII* codifica per una proteina che protegge CII dalla proteolisi, competendo per la proteasi, essendo anch'essa bersaglio di FtsH. CII è usata da lambda come un vero e proprio sensore dello stato di salute della cellula ospite.

CII attiva anche altri due promotori necessari alla crescita lisogenica, P_i e P_{AQ} , dei quali il primo regola la trascrizione del gene *int*, che serve a promuovere la ricombinazione sito-specifica tra i siti di lambda e di *E. coli* per permettere l'integrazione.

Oltre alle condizioni di crescita, anche il numero di particelle fagiche che infettano una singola cellula di *E. coli* è importante per la scelta tra i due cicli vitali di lambda. Questo numero, definito **molteplicità di infezione**, indica quanti fagi infettano una cellula batterica all'interno di una popolazione. Quando una cellula batterica è infettata in media da una sola particella di fago, si instaurerà con molta probabilità il ciclo litico. Se invece il batterio è infettato da più di due particelle contemporaneamente, è probabile che si andrà verso la lisogenia. Questo fenomeno è sempre legato all'interruttore CII; infatti, più fagi infettano un singolo batterio, più trascrizione ci sarà dai promotori P_L e P_R e quindi i livelli di CIII e CII saranno più elevati e in grado di attivare sia la trascrizione del repressore che della proteina Int. È sufficiente che questo avvenga in un solo cromosoma di lambda perché la cellula divenga immune all'instaurazione del ciclo litico, proprio grazie all'efficienza del repressore a usare il legame cooperativo per spegnere la trascrizione di tutti i geni di lambda. Infatti, il legame del repressore ai siti di controllo inibisce la trascrizione di N e quindi di tutti i geni del fago, eccetto se stesso, dal promotore P_{RM} . In un certo senso si può dire che il repressore agisce come regolatore "epigenetico" in quanto riesce a mantenere il ciclo lisogenico per generazioni e a proteggere il batterio da nuove infezioni.

Controllo dell'integrazione e dell'escissione del fago nel cromosoma di *Escherichia coli*

L'integrazione del DNA del fago nel cromosoma batterico dipende da una sola proteina che si chiama **Int (integrasi)**, mentre la reazione opposta, vale a dire l'escissione, dipende dall'azione congiunta di Int e di **Xis (escissionasi)**. Gli aspetti molecolari dell'integrazione e dell'escissione sono stati trattati nel paragrafo 8.2.

Questi due geni sono contigui e vari meccanismi di controllo fanno sì che siano attivi solo quando la loro funzione è richiesta. Ci sono diverse situazioni in cui questi geni devono essere diversamente regolati: durante il ciclo lisogenico, durante il ciclo litico e quando un profago, integrato nel cromosoma dell'ospite, deve rispondere all'induzione causata da raggi UV e, quindi, essere escisso per intraprendere il ciclo litico. Descriviamo qui le tre possibili situazioni.

1. Nel primo caso, quando va verso il ciclo lisogenico, per integrare il suo genoma in quello della cellula ospite il fago deve produrre livelli elevati di Int e livelli quanto più possibile bassi di Xis. In **figura 10.42A** si vede che la proteina CII, agendo sul promotore P_i , attiva la trascrizione che poi si arresta a un sito di terminazione a valle di *int*. L'mRNA sintetizzato contiene al 5' solo parte della sequenza codificante di *xis* (che quindi non viene tradotta) a cui segue l'intera sequenza codificante di *int* che risulta in un'alta produzione di Int. Va detto che una certa quantità di Int e Xis possono essere prodotte dall'mRNA sintetizzato a partire dal promotore P_L , ma l'efficienza del promotore P_i rende comunque i livelli di Int molto più alti di Xis.
2. Nel caso della crescita litica, il fago deve produrre livelli alti di Xis per evitare che il cromosoma di lambda in replicazione venga integrato nel cromosoma dell'ospite. In questa via di sviluppo si suppone che CII non sia molto attivo e quindi il messaggero preponderante sarà quello prodotto dal promotore P_L , che comprende le sequenze codificanti sia per Xis che per Int. Per effetto di pN che agisce da antiterminatore, la trascrizione che parte da questo promotore continua oltre i terminatori di *N* e di *int* e la polimerasi trascrive in un unico mRNA anche la regione Sib, come illustrato in **figura 10.42B**. Questa molecola di RNA va incontro a una degradazione sequenziale da parte della RNasi III, che riconosce una forcina che si forma nella regione di *sib*. Come si può vedere in **figura 10.42B** la degradazione comincia dall'estremità più vicina a Int e quindi questa regione viene degradata prima di *xis* e questo mRNA produce quindi più Xis che Int. Tale fenomeno viene definito **retro-regolazione** ed è un controllo post-trascrizionale.

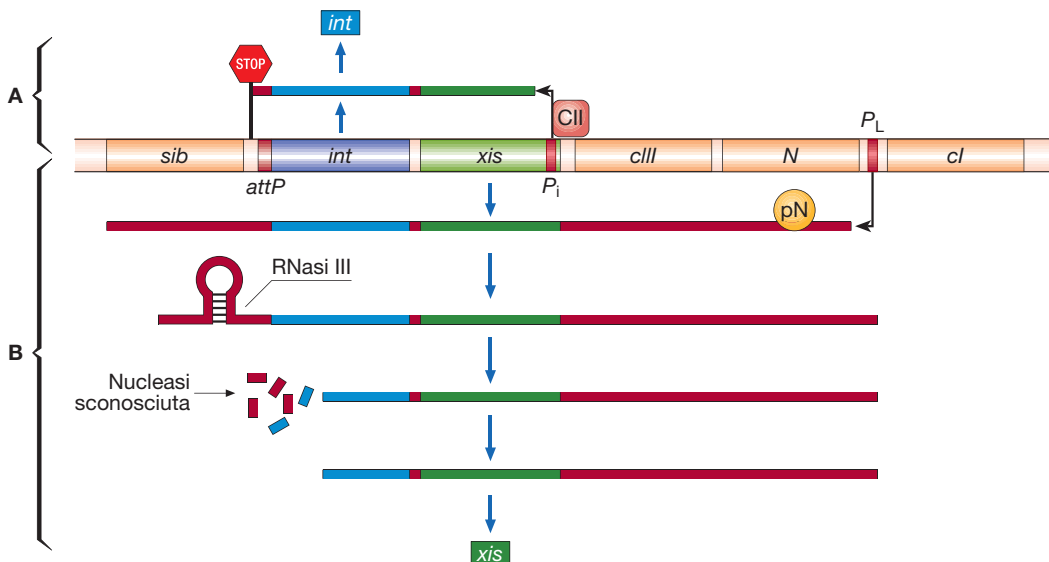


Figura 10.42 L'entrata nel ciclo lisogenico comporta una preponderante produzione di Int, mentre l'entrata nel ciclo litico comporta una preponderante produzione di Xis. (A) Nel caso del ciclo lisogenico, la proteina CII attiva la trascrizione di Int agendo sul promotore P_i , con produzione della proteina Int ma non della proteina Xis, in quanto il trascritto manca della sequenza iniziale

necessaria per la traduzione della prima ORF. (B) Nel ciclo litico il trascritto preponderante inizia dal promotore P_L e prosegue fino alla fine della regione *sib*; questa contiene una struttura a forcina che innesca una degradazione sequenziale della regione *sib* e *int* dell'mRNA. Xis resta, in questo caso, il principale prodotto proteico sintetizzato.

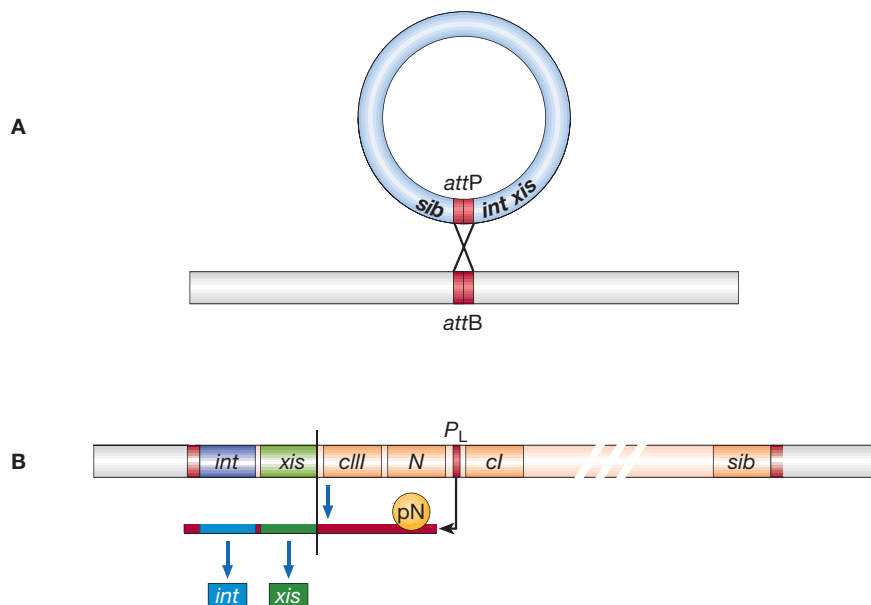


Figura 10.43 L'escissione del genoma fagico dal cromosoma dell'ospite durante l'induzione con radiazioni UV comporta alti livelli di espressione sia di Int che di Xis. Il processo d'integrazione porta a una separazione della regione *sib* dalla sequenza codificante di *int*. In questo modo non può essere innescata la degradazione sequenziale dell'mRNA, con il risultato di un'alta produzione sia di Int che di Xis.

3. Quando, dopo esposizione a radiazioni UV di un batterio lisogenico, viene indotta la fase litica, il fago deve produrre grandi quantità sia di Int che di Xis per poter fuggire da una cellula danneggiata. Poiché il promotore P_i è inattivo per la mancanza di CII in questa fase di crescita, viene usato il promotore P_L , che produce un mRNA contenente la regione codificante sia per Xis che per Int. In questo caso l'RNA non contiene la sequenza *sib*. Infatti, se si osserva la struttura del genoma di lambda integrato in *E. coli*, come illustrato in **figura 10.43A**, si può vedere che il processo di integrazione di lambda nel cromosoma di *E. coli*, mediante i siti *attP* e *attB*, separa Sib da Int in modo tale che l'mRNA prodotto sia stabile e produca livelli elevati delle proteine Xis e Int, necessarie alla escissione per garantirsi la fuga dalle cellule irradiate.

La regolazione di lambda, come abbiamo visto in questo capitolo, si basa su diversi meccanismi che, in maniera molto fine, controllano le varie fasi di sviluppo di questo batteriofago. Alcuni di essi, come la dipendenza della crescita dalle condizioni ambientali, l'attivazione trascrizionale e i controlli post-trascrizionali che degradano l'RNA, rappresentano un modello, seppur complicato, sicuramente molto più semplice dei processi alla base della regolazione dei geni degli organismi eucariotici.

Lecture di approfondimento

- Jacob, F. e Monod, J. (1961), «Genetic regulatory mechanisms in the synthesis of proteins», *Journal of Molecular Biology*, **3**, pp. 318-356.
- Lawson, C.L. *et al.* (2004), «Catabolite activator protein: DNA binding and transcription activation», *Current Opinion in Structural Biology*, **14**(1), pp. 10-20.
- Ptashne, M. (2005), «Regulation of transcription: from lambda to eukaryotes», *Trends in Biochemical Sciences*, **30**(6), pp. 275-279.
- Ptashne, M. (2006), «Lambda's switch: lessons from a module swap», *Current Biology*, **16**(12), pp. 459-462.
- Ptashne, M., *Regolazione genica*, Zanichelli, Bologna, 2004.

11

Trascrizione e regolazione negli eucarioti

11.1 Tre sistemi di trascrizione nel nucleo

11.2 Promotore dei geni per gli rRNA e fattori che regolano la trascrizione dell'RNA polimerasi I

11.3 RNA polimerasi III: promotori di Pol III interni ed esterni e suoi fattori di trascrizione

11.4 RNA polimerasi II: struttura del promotore minimo di Pol II

11.5 Regolazione

11.6 Struttura modulare e domini dei fattori di trascrizione (transattivatori)

11.7 Domini funzionali dei transattivatori: domini di legame al DNA

Come abbiamo visto nel capitolo 10, i procarioti hanno una sola RNA polimerasi; negli eucarioti, invece, esistono almeno tre differenti RNA polimerasi, oltre a quelle che troviamo negli organelli, che trascrivono classi specifiche di geni.

L'**RNA polimerasi I (RNA Pol I)** trascrive i geni per l'rRNA, l'**RNA polimerasi II (RNA Pol II)** trascrive i precursori degli mRNA e piccoli RNA implicati, a vari livelli, nella regolazione post-trascrizionale (*vedi* par. 12.4), mentre l'**RNA polimerasi III (RNA Pol III)** trascrive i geni per l'rRNA 5S, per i tRNA e per piccoli RNA coinvolti nella formazione di molecole di mRNA mature. (Attenzione a non confondere queste tre RNA polimerasi eucariotiche con le tre DNA polimerasi – I, II e III – procariotiche di cui abbiamo parlato nel capitolo 6.) Ognuna delle tre RNA polimerasi funziona in associazione con una serie di fattori specifici che riconoscono elementi in *cis* sul DNA nella regione del promotore. Mentre i promotori dell'RNA polimerasi I e dell'RNA polimerasi III sono abbastanza semplici, con gli elementi del promotore nelle vicinanze del sito d'inizio della trascrizione, nel caso dell'RNA polimerasi II la regione del promotore può essere molto estesa con un grado di complessità che riflette la diversità dei geni trascritti, così come la necessità di modulare la loro trascrizione in senso spazio-temporale e in risposta a vari stimoli.

Negli eucarioti l'espressione dei geni è regolata a vari livelli e, certamente, il controllo a livello trascrizionale è essenziale e fondamentale per la vita di ogni cellula. Va, inoltre, sottolineato che gli organismi multicellulari sono costituiti da una varietà di tipi differenti di cellule e tessuti in grado di svolgere funzioni precise nell'ambito dell'intero organismo. Ciascuna di queste cellule nell'uomo, pur svolgendo specifiche funzioni, contiene gli stessi geni (circa 25 000) che costituiscono il suo genoma (*vedi* par. 5.3). La straordinaria diversità nella specializ-

zazione cellulare è ottenuta attraverso l'accurato controllo dell'espressione e della regolazione di vari geni; inoltre, la risposta di ogni cellula a stimoli fisiologici e ambientali richiede un accurato controllo della risposta a questi stimoli.

Molti meccanismi che abbiamo incontrato nella regolazione della trascrizione dei batteri sono utilizzati anche per la regolazione a livello trascrizionale negli eucarioti; tuttavia, come atteso, la regolazione della trascrizione negli eucarioti è più complessa, proprio per le molteplici funzioni e specializzazioni che ciascuna cellula eucariotica deve poter realizzare. Anche negli eucarioti, dunque, la trascrizione è regolata da attivatori e repressori, ma ci sono ulteriori livelli di complessità che sono assenti nei procarioti. Uno di questi è rappresentato dall'organizzazione del DNA in una struttura compatta, nota come cromatina, che abbiamo già descritto nel paragrafo 4.8. L'impacchettamento del DNA in tale struttura rappresenta un ostacolo per l'RNA polimerasi e per i fattori di trascrizione che devono riconoscere e legare specifiche sequenze di DNA, quale un promotore. Nei procarioti il DNA è per lo più facilmente disponibile all'interazione con l'RNA polimerasi e proteine che agiscono come repressori o attivatori della trascrizione; negli eucarioti, invece, questa disponibilità coinvolge, come vedremo meglio in seguito, un raffinato sistema di regolazione. La modulazione del grado di compattezza della cromatina è, quindi, essenziale per rendere inaccessibili o disponibili sequenze di DNA con cui interagiscono le RNA polimerasi eucariotiche e le proteine regolative che ne influenzano l'attività. Questo livello di regolazione sarà trattato alla fine di questo capitolo e nel paragrafo 17.2.

Un'altra importante differenza tra la trascrizione nei procarioti e negli eucarioti riguarda il ruolo che i fattori di trascrizione hanno nel reclutare le RNA polimerasi sui promotori. Nei batteri, la polimerasi si lega al promotore riconoscendo, attraverso la subunità σ , brevi sequenze consenso che si trovano, generalmente, molto vicine al sito d'inizio della trascrizione; solo in alcuni casi le sequenze del promotore non sono sufficienti a far legare stabilmente la polimerasi. Negli eucarioti, invece, l'RNA polimerasi non è praticamente in grado di legarsi da sola al promotore e il ruolo dei fattori di trascrizione è essenziale per il suo reclutamento sul DNA. Questi fattori, come vedremo, possono legarsi a brevi sequenze del DNA e richiamare sulla regione che controllano e che deve essere trascritta complessi enzimatici che rimodellano e aprono la cromatina permettendo ai fattori basali e al mediatore (*vedi oltre*) di legarsi al promotore e richiamare la polimerasi. Un'altra differenza fondamentale tra eucarioti e procarioti è che mentre nei batteri i geni sono organizzati in operoni e trascritti come mRNA policistronici, negli eucarioti ogni gene è regolato e trascritto singolarmente. Con le nuove metodologie della trascrittomica (*vedi par. 21.2*) si è rivelato un quadro molto più complesso di quello che si prevedeva per la popolazione generale di RNA trascritti, includendone molti che non producono proteine. Una delle sorprese più recenti è rappresentata da una nuova classe di trascritti che partono vicino ai siti di inizio a monte delle sequenze che codificano per le proteine. Tuttavia, questi RNA sono corti, presenti in quantità minima e spesso trascritti in direzione opposta a quella del gene codificante. Non si conosce la loro funzione, ma la loro esistenza rende più complessa la nostra visione di come le sequenze di DNA conosciute come "promotori" definiscono i siti di inizio della trascrizione (TSS, *Transcription Start Site*).

11.1 Tre sistemi di trascrizione nel nucleo

Come accennato sopra, le tre RNA polimerasi eucariotiche sono specializzate nel trascrivere classi diverse di geni. **Pol I**, trascrivendo gli rRNA, è localizzata principalmente nel nucleolo. **Pol II** è invece localizzata nel nucleoplasma, dove trascrive i precursori degli mRNA, i miRNA (piccoli RNA che regolano l'espressione genica attraverso il meccanismo dell'interferenza), e alcuni snRNA

che hanno un ruolo nel processamento degli mRNA. **Pol III**, pure nucleoplasmatica, trascrive i geni che codificano per i tRNA, l'rRNA 5S, l'snRNA U6 coinvolto nello splicing e l'RNA 7SL parte del complesso SRP, che ha un ruolo nella regolazione della traduzione. Nelle piante esistono altre due RNA polimerasi non essenziali per la vita delle cellule, **Pol IVa** e **Pol IVb**, che hanno un ruolo importante in diversi aspetti del controllo del silenziamento di geni. Le RNA polimerasi eucariotiche sono facilmente distinguibili sperimentalmente sulla base della loro sensibilità all' α -amanitina, una potente tossina prodotta dal fungo *Amanita phalloides*. L'RNA polimerasi più sensibile è Pol II, mentre la più resistente è Pol I.

Le tre diverse RNA polimerasi eucariotiche e la polimerasi di *Escherichia coli* sono messe a confronto nella **figura 11.1**. Le polimerasi eucariotiche sono costituite da circa 12 subunità che, nel loro insieme, formano un complesso proteico con una dimensione superiore ai 500 kDa. Tra queste subunità ne distinguiamo alcune (1, 2, 3, 11 e 6) che sono omologhe, rispettivamente, alle subunità β' , β , $\alpha 1$, $\alpha 2$ e ω dell'enzima procariotico; si tratta quindi di subunità presenti in tutte le RNA polimerasi procariotiche ed eucariotiche. I tre enzimi eucariotici hanno inoltre in comune alcune subunità (5, 8, 10 e 12) che mancano invece nella polimerasi procariotica. Infine, un certo numero di altre subunità sono specifiche per ciascuno dei tre enzimi eucariotici. Va notato che quanto detto è molto semplificato e approssimato, in parte perché è difficile definire due subunità come uguali od omologhe o simili. Inoltre, abbiamo preso in considerazione solo l'RNA polimerasi di *E. coli* e le tre principali RNA polimerasi eucariotiche, tralasciando di includere nel confronto le RNA polimerasi degli archea, degli organelli e degli eucarioti di diversa complessità (per es. lievito e mammiferi).

La **figura 11.1** mostra anche una caratteristica che contraddistingue Pol II rispetto alle altre polimerasi, cioè la presenza del CTD (*C-Terminal Domain*,

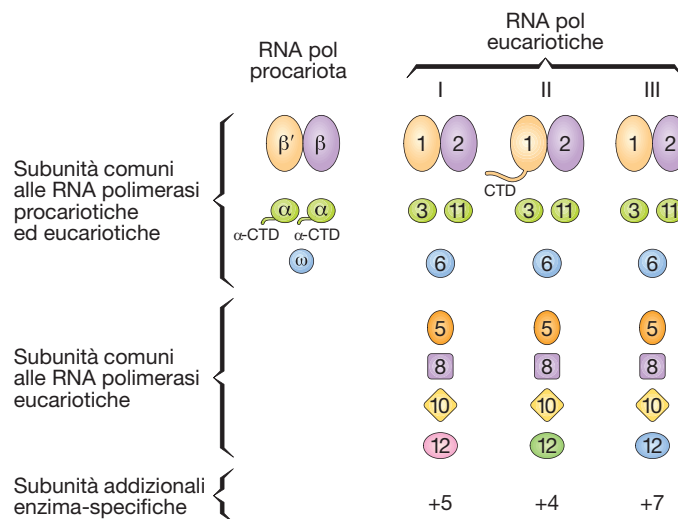


Figura 11.1 Confronto tra la costituzione in subunità delle tre RNA polimerasi degli eucarioti e l'RNA polimerasi di *Escherichia coli*. Le RNA polimerasi eucariotiche sono costituite da circa 12 subunità. Cinque di queste (Rpb1, Rpb2, Rpb3, Rpb11 e Rpb6) sono omologhe, rispettivamente, alle subunità β' e β , alle due subunità α e alla subunità ω della polimerasi procariotica. Inoltre le tre polimerasi eucariotiche contengono alcune subunità in comune (Rpb5, Rpb8, Rpb10 e Rpb12) che sono invece assenti nell'enzima procariotico. Infine, le tre polimerasi eucariotiche contengono alcune subunità specifiche per ciascuna di esse. Si noti che, per

semplicità, nella figura è riportata solo la numerazione delle subunità eucariotiche omettendo il prefisso Rpb. La subunità Rpb1 della polimerasi II eucariotica presenta il CTD (dominio carbossi-terminale) costituito da una ripetizione della sequenza amminoacidica -Tyr-Ser-Pro-Thr-Ser-Pro-Ser-. Questo dominio è assente nella polimerasi procariotica, dove invece è presente un dominio C-terminale della subunità α , chiamato α -CTD, coinvolto nel riconoscimento delle sequenze a monte di alcuni promotori (vedi par. 9.4).

dominio C-terminale) della subunità Rpb1. Questo CTD consiste di una lunga coda C-terminale non strutturata costituita da ripetizioni della sequenza Tyr-Ser-Pro-Thr-Ser-Pro-Ser. Il CTD, che è diverso per struttura e funzione dall' α -CTD della RNA polimerasi procariotica, ha un ruolo fondamentale nella formazione del complesso d'inizio di trascrizione e nel distacco dell'enzima dal promotore; la funzione del CTD è regolata dallo stato di fosforilazione dei suoi residui. Nel lievito *Saccharomyces cerevisiae* ci sono circa 26 ripetizioni della sequenza amminoacidica sopra indicata, che nei mammiferi può essere presente anche in 52 copie. Il numero delle ripetizioni è importante: infatti, mutazioni che fanno perdere più della metà di queste sequenze sono letali in vari sistemi biologici.

La somiglianza strutturale delle diverse RNA polimerasi eucariotiche fa pensare che l'origine di questi enzimi sia probabilmente dovuta alla duplicazione di geni codificanti per subunità ancestrali, che si sono poi specializzate nelle funzioni di ciascun enzima e, soprattutto, nell'essere in grado di interagire specificamente con i vari fattori che ne controllano l'attività. Pol II deve essere in grado di trascrivere un numero elevato di geni diversi, mentre Pol I deve produrre in quantità abbondante una sola specie di trascritto e Pol III molte copie dei tRNA e del 5S RNA. Tuttavia, la funzione catalitica di tutti questi enzimi, cioè l'aggiunta progressiva di nucleotidi a catene crescenti di RNA, è un meccanismo molto conservato deducibile dal grado di conservazione e somiglianza del sito attivo di tutte le RNA polimerasi sia procariotiche che eucariotiche.

Abbiamo già accennato che le RNA polimerasi eucariotiche non sono in grado, da sole, di riconoscere e legare i loro promotori. Hanno bisogno di fattori addizionali che permettono il riconoscimento degli elementi in *cis* del promotore e il successivo reclutamento dei rispettivi enzimi. Questo concetto, infatti, è valido per tutte e tre le RNA polimerasi: esistono, quindi, fattori specifici per ciascuna di esse. Nel caso di Pol I e Pol III il numero di fattori specifici per il reclutamento sui relativi promotori è abbastanza limitato. Nel caso di Pol II il quadro è più complesso. Infatti, esistono essenzialmente due famiglie di fattori di trascrizione: la prima, costituita dai "fattori basali" di trascrizione, è richiesta per la formazione del complesso d'inizio sui promotori dei geni trascritti da Pol II; la seconda è costituita da proteine che sono responsabili della regolazione della trascrizione, sia come attivatori che come repressori. Questa complessità è evidenziata dalla complicata struttura dei promotori di Pol II, che contengono elementi in *cis* che mediano il legame specifico di entrambe le famiglie di fattori trascrizionali. La **figura 11.2** mette a confronto schematicamente il promotore procariotico e i promotori relativi alle tre RNA polimerasi eucariotiche I, II e III.

11.2 Promotore dei geni per gli rRNA e fattori che regolano la trascrizione dell'RNA polimerasi I

L'RNA polimerasi I trascrive esclusivamente i geni per gli RNA ribosomali. Negli eucarioti questi geni sono presenti in copie multiple (generalmente decine o centinaia), tutte uguali tra loro, organizzate in grandi cluster di operoni ripetuti in tandem, localizzati su uno o più cromosomi (*vedi* par. 12.1 e fig. 12.2). Questi geni sono separati da spaziatori che contengono, ciascuno, il promotore del gene a valle. A partire da questo promotore viene trascritta una molecola di RNA precursore che contiene i grandi e piccoli rRNA che sono successivamente separati tra di loro.

Il promotore del precursore degli rRNA è illustrato schematicamente in **figura 11.2B**. Esso è costituito da due regioni: un tratto che si sovrappone al sito d'inizio della trascrizione TSS (si estende dalle posizioni -45 a +20) che costituisce il nucleo del promotore (core) e un elemento più a monte, chiamato **elemento UPE** o **UCE** (*Upstream Promoter Element* o *Upstream Control Element*), che si

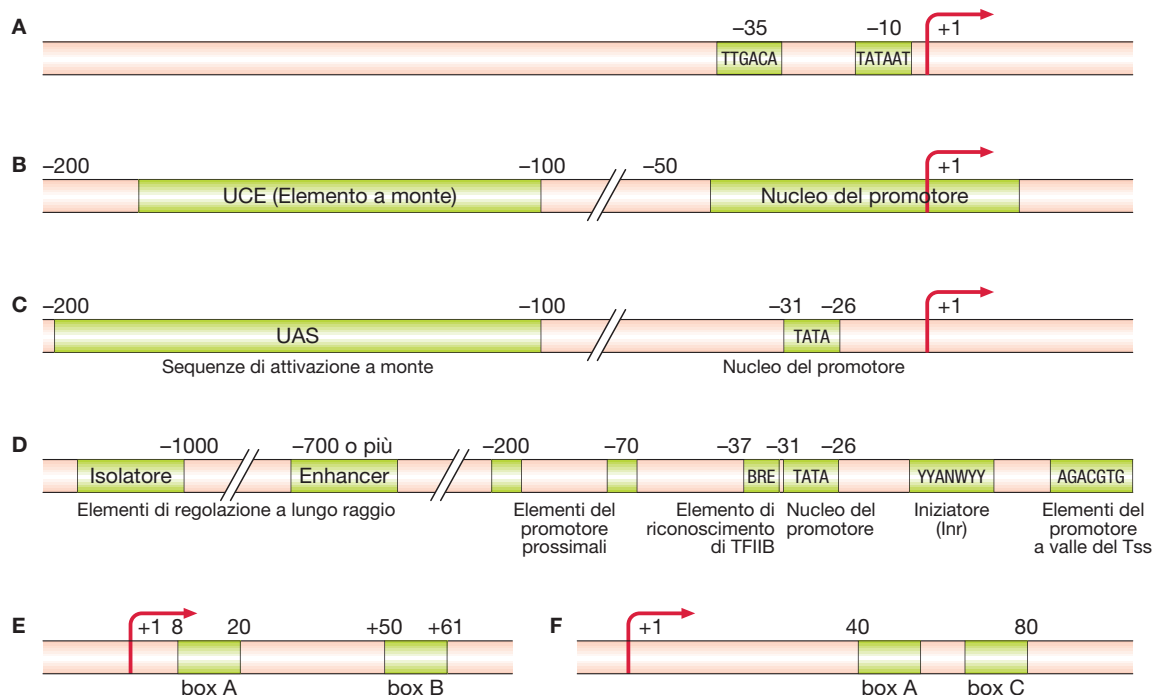


Figura 11.2 Promotore batterico e promotori delle tre polimerasi eucariotiche a confronto. (A) Promotore dei batteri dove sono presenti le sequenze -10 e -35, riconosciute dal fattore sigma. (B) Promotore di Pol I negli eucarioti. È costituito da due regioni: un tratto sovrapposto al sito d'inizio della trascrizione (TSS, indicato dalla freccia rossa) costituisce il nucleo del promotore (core); un elemento più a monte, chiamato elemento UPE o UCE, che si estende dal nucleotide -107 fino a -180. (C) e (D) Promotori di Pol II negli eucarioti. Nel caso degli eucarioti unicellulari (C) il promotore è tipicamente costituito da una sequenza TATA che si trova a -26 dal TSS e da un elemento UAS (*Upstream Activating Sequence*) che viene riconosciuto dagli attivatori specifici per quel trascritto. Negli eucarioti multicellulari (D) i promotori sono più estesi e complessi dei corrispondenti di lievito, e sono anche molto eterogenei. Essi contengono sequenze di riconoscimento, nella parte vicina (prossimale) al TSS, costituite in genere dalla TATA box, in posizione -26, e da una serie di elementi posizionati sia

a monte (fino a -200 pb) che a valle del TSS. La TATA box è presente in circa il 50% dei promotori. A questi elementi vicini al sito d'inizio della trascrizione (*proximal elements*) si legano in genere i fattori basali. Sono presenti poi elementi distanti, gli enhancer, dove si legano gli attivatori specifici per il gene da trascrivere, che possono trovarsi anche a distanze di varie kpb dal TSS. Oltre agli elementi enhancer, sono presenti elementi definiti isolatori (*insulators*) che regolano l'azione degli enhancer annullando la loro funzione. Non tutti i geni trascritti dalla RNA polimerasi II contengono tutti questi elementi. (E) e (F) Promotori dei geni per i tRNA (E) e per l'rRNA 5S (F) trascritti da Pol III negli eucarioti. Gli elementi del promotore sono in questo caso localizzati nella regione trascritta del gene, a valle del TSS. Nel caso dei geni per i tRNA i promotori sono costituiti da una box A e una box B, mentre nel caso dei geni per l'rRNA 5S sono presenti una box A e una box C. Questi elementi sono riconosciuti dai fattori d'inizio che posizionano la polimerasi a monte del TSS.

estende dal nucleotide -107 fino a -180. Il core è ricco di basi G-C nella regione tra -40 e -20, mentre intorno alla regione d'inizio è ricco di basi A-T. Anche l'elemento UPE è ricco di basi G-C.

La formazione del complesso d'inizio, illustrata in **figura 11.3**, richiede almeno due fattori trascrizionali ausiliari, chiamati **Upstream Binding Factor (UBF)** e **SL1**. UBF è una proteina che si lega al DNA con domini multipli del tipo **HMG (High Mobility Group)** e interagisce direttamente con il core del promotore e con l'elemento UCE. La regione C-terminale di UBF è necessaria per l'attivazione della trascrizione ed è ampiamente fosforilata. SL1 è un complesso multiproteico che comprende la **TATA Binding Protein (TBP)**, che è un componente essenziale dei complessi d'inizio di tutte e tre le RNA polimerasi eucariotiche, e altri fattori associati (**TAF, TBP Associated Factors**) specifici per Pol I. SL1 non si lega al DNA riconoscendo una sequenza specifica, ma il suo reclutamento sul promotore è mediato da interazioni specifiche con UBF. Una volta posizionato sul promotore, il complesso SL1 contatta direttamente il DNA e promuove la trascrizione, reclutando Pol I. Il dominio C-terminale di UBF e il suo stato di fosforilazione influenzano il legame con SL1.

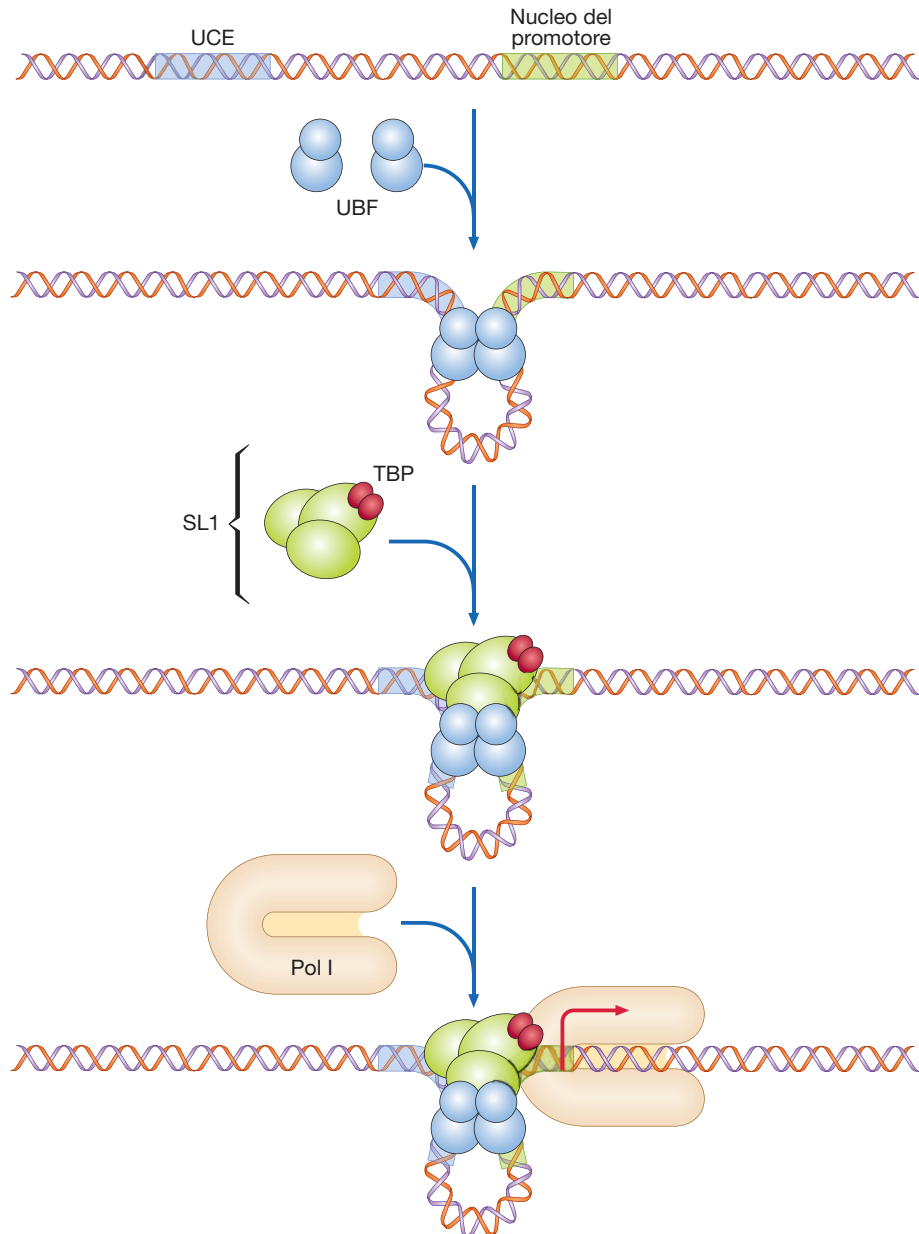


Figura 11.3 Formazione del complesso di pre-inizio della RNA polimerasi I. Il fattore UBF (*Upstream Binding Factor*) si lega agli elementi del nucleo del promotore e alla regione UCE (*Upstream Control Element*) in forma di dimero. Una volta legato, UBF richiama il complesso SL1, che contiene la TBP e tre TAF (*TBP-Associated Factors*) specifici per la trascrizione di Pol I. Una volta formato, il complesso richiama l'RNA polimerasi I sul sito d'inizio della trascrizione.

Come mostrato in **figura 11.3**, un dimero di UBF si lega sia a UCE che al core del promotore, promuovendo la formazione di un'ansa sul DNA. A questo punto interviene SL1 che richiama Pol I sul promotore e la sintesi dell'rRNA può iniziare.

Come già detto sopra, le cellule somatiche degli eucarioti superiori hanno circa 200 copie dei geni per l'rRNA per genoma aploide e solo una frazione di esse è trascritta in un certo intervallo temporale, mentre le altre copie sono silenziate. La regolazione della trascrizione degli rRNA comporta, quindi, il controllo del numero di geni che sono trascritti e la frequenza con la quale ciò avviene. Si pensa che UBF abbia un ruolo essenziale in questo aspetto regolativo.

11.3 RNA polimerasi III: promotori di Pol III interni ed esterni e suoi fattori di trascrizione

L'RNA polimerasi III trascrive i geni per i tRNA e per l'rRNA 5S utilizzando promotori interni (che si sovrappongono, cioè, alle regioni trascritte) e che contengono degli elementi (**box**) di controllo (**fig. 11.2E e F**). Esiste anche un terzo tipo di promotore per Pol III, non mostrato nella figura, che promuove la trascrizione dei piccoli RNA nucleari (snRNA) e che contiene elementi di controllo, tra cui un tipico elemento TATA, a monte del sito d'inizio della trascrizione (*vedi* oltre). Oltre a questa sequenza, che permette il legame specifico di TBP, il promotore per gli snRNA può contenere anche gli elementi Oct e PSE, che sono presenti anche nei promotori degli snRNA trascritti dalla RNA polimerasi II. La presenza di questi elementi sul promotore aumenta l'efficienza della trascrizione di Pol III.

Come funzionano i promotori che si trovano a valle del sito d'inizio della trascrizione (TSS)? Come mostrato in **figura 11.4**, nel caso dei tRNA il fattore TFIIC si lega agli elementi interni denominati box A e box B. Questo legame, che è indipendente da altri fattori, richiama sul promotore TFIIB. Questo è composto da diverse subunità tra le quali TBP, che abbiamo già visto essere presente nel fattore SL1 di Pol I. La formazione di questo complesso richiama la RNA polimerasi III sul TSS. Nel caso del gene per l'rRNA 5S, il fattore specifico TFIIA, che è costituito da domini di legame al DNA del tipo **dita di zinco** (*vedi* oltre), si lega specificamente alla box A. TFIIA permette il posizionamento di TFIIC sul promotore, che a sua volta richiama TFIIB sul sito d'inizio della trascrizione. Come mostrato in figura, TFIIC e TFIIA sono dei fattori di assemblaggio il cui compito è di posizionare in modo corretto TFIIB sul TSS. La funzione della componente TBP, che fa parte di TFIIB, è di reclutare Pol III precisamente sul sito d'inizio, analogamente a quanto accade nel caso della RNA polimerasi II. TBP si lega al DNA e lo piega di 80°: tale curvatura del DNA è stata osservata sia per i geni trascritti da Pol III che da Pol II. TFIIB è, quindi, un fattore di posizionamento che funziona nello stesso modo su tutti i tipi di promotori utilizzati da Pol III. Nel caso dei promotori con sequenze di controllo a monte del sito d'inizio, TFIIB è in grado di legarsi da solo, in quanto riconosce l'elemento TATA, posto vicino al sito d'inizio della trascrizione.

11.4 RNA polimerasi II: struttura del promotore minimo di Pol II

Come abbiamo già detto, la situazione per l'RNA polimerasi II è molto più complessa rispetto a Pol I e Pol III. Innanzitutto il promotore, cioè la regione di DNA coinvolta nell'inizio della trascrizione e nella regolazione, è in genere molto più lungo; in secondo luogo, il numero di fattori di trascrizione richiesti è molto più elevato. Esperimenti preliminari avevano stabilito che Pol II ha bisogno, per iniziare una efficiente trascrizione *in vitro*, di un estratto cellulare. Il frazionamento dei vari componenti di questo estratto ha portato all'identificazione essenzialmente di due gruppi di fattori di trascrizione. Il primo gruppo, costituito da una serie di proteine chiamate **fattori basali** (chiamati anche **generali**, da cui l'acronimo **GTF**) è richiesto per reclutare l'RNA polimerasi su tutti i promotori di Pol II e formare il complesso d'inizio; l'RNA polimerasi II associata a questi fattori costituisce l'**apparato basale** della trascrizione di Pol II. Il secondo gruppo di fattori è costituito da proteine che sono responsabili della regolazione della trascrizione, sia come attivatori che come repressori. La struttura dei promotori di Pol II è piuttosto complessa, è molto differente per i diversi geni e per i diversi organismi eucariotici, e comprende varie importanti

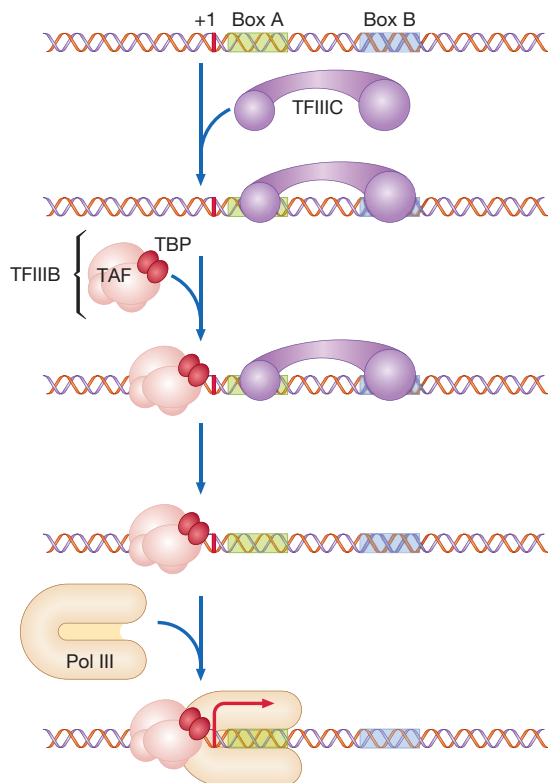
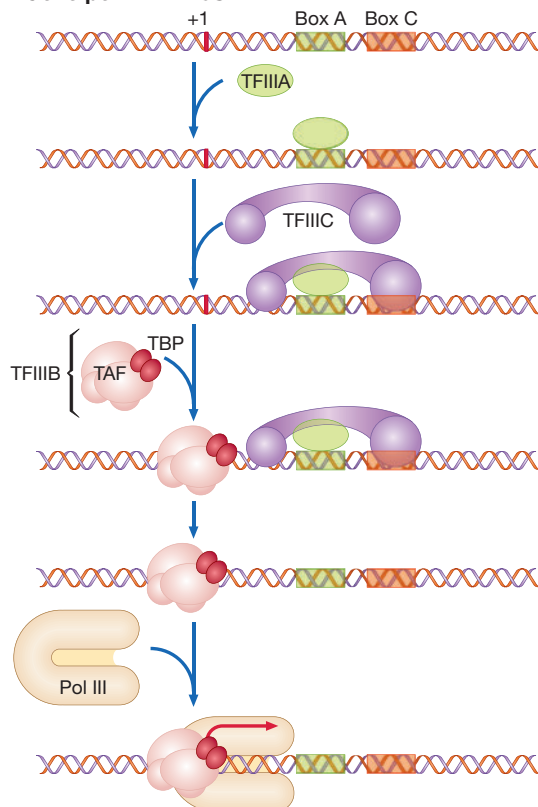
A. Gene del tRNA**B. Gene per l'rRNA 5S**

Figura 11.4 Formazione del complesso di pre-inizio della RNA polimerasi III. La formazione del complesso di pre-inizio di Pol III richiede il riconoscimento di sequenze all'interno della regione trascritta. (A) Nel caso dei geni per i tRNA, solo il fattore TFIIIC è necessario per il riconoscimento specifico. Esso si lega contemporaneamente alle box A e B e richiama il fattore TFIIIB che contiene TBP posizionandolo a monte del sito d'inizio. Il complesso così formato recluta la polimerasi sul promotore. (B) Nel caso del gene per il 5S, il legame del fattore specifico TFIIIA alla box A permette il successivo legame del fattore TFIIIC alla box C. Questo poi richiama TFIIIB posizionandolo a monte del promotore e permettendo l'inizio della trascrizione.

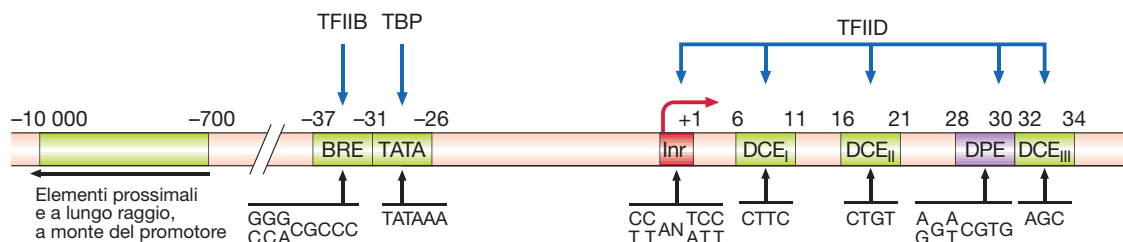


Figura 11.5 Promotore “minimo” per l’RNA polimerasi II. Sono qui riportati gli elementi che costituiscono i promotori minimi per l’RNA polimerasi II; non tutti gli elementi sono sempre presenti. L’elemento Inr (in rosso) contiene, indicato dalla freccia rossa, il sito d’inizio della trascrizione TSS (*Transcription Start Site*), che è in genere una A fiancheggiata da alcune pirimidine. La TATA box (in verde), che rappresenta l’elemento più importante e più frequente, è localizzata a circa 26 nt a monte del TSS ed è spesso preceduta dall’elemento BRE e alle volte è accompagnata

da una o più DCE box localizzate a valle del TSS (tutte in verde). L’elemento DPE (in violetto) è caratteristico dei promotori “TATA-less”, cioè privi della TATA box. Le sequenze consenso, che corrispondono ai vari elementi del promotore, sono riportate sotto. Si noti che le sequenze riportate si riferiscono al filamento codificante della doppia elica e che, in alcune posizioni, si possono trovare nucleotidi diversi. In alto sono indicati i nomi dei fattori basali di trascrizione che interagiscono con i vari elementi sul DNA.

sequenze che operano in *cis*. Uno schema generale della struttura dei promotori di Pol II è mostrato in **figura 11.2C** per gli eucarioti unicellulari (*Saccharomyces cerevisiae*) e in **figura 11.2D** per gli eucarioti multicellulari.

La **figura 11.5**, invece, mostra quello che viene chiamato il **promotore minimo** (**core promoter**) di Pol II che è costituito dal tratto di DNA indispensabile per permettere ai fattori basali di far iniziare la trascrizione. Tale promotore minimo contiene l’elemento chiamato **iniziatore** o **INR** e che ha una sequenza conservata YYANWYY (*vedi* tab. 21.2 per i codici IUB delle basi ambigue), da cui parte la trascrizione iniziando su un nucleotide, generalmente una A, che abbiamo già chiamato TSS. Un altro elemento molto importante è una sequenza posizionata a circa -26 dal sito d’inizio e chiamata **TATA box** (sequenza conservata TATAAA). Tale sequenza è, in genere, preceduta da un altro elemento chiamato **BRE** (*B Responsive Element*) che permette il legame del fattore basale TFIIB. A valle del TSS vi possono essere altri elementi che caratterizzano un promotore (**fig. 11.5**). Un elemento chiamato **DPE** (*Downstream Promoter Element*) è posto tra le posizioni +28 e +34 ed è tipicamente presente in quei promotori che non contengono la TATA box, chiamati **TATA-less**. Invece l’elemento **DCE** (*Downstream Core Element*) può essere presente in più copie (chiamate DCE_I, DCE_{II} e DCE_{III}) nei promotori contenenti la TATA box. Altri elementi che sono presenti nel promotore sono le **isole CpG**. Questi elementi sono costituiti da dinucleotidi GC. Più del 50% dei promotori umani contiene queste sequenze. Solo una frazione dei promotori associati alle CpG contengono anche la TATA box. Delle isole CpG parleremo in seguito nel capitolo 17 (*vedi* par. 17.1), quando descriveremo il ruolo della metilazione del DNA nel controllo dell’espressione genica.

Come illustrato in **figura 11.2**, oltre agli elementi che costituiscono il promotore minimo, la struttura generale di un promotore di Pol II contiene elementi localizzati fino a 200 pb a monte del sito d’inizio (che vengono definiti **elementi prossimali**) ed elementi più distali, chiamati **enhancer**, che possono essere posizionati fino a 10 kpb dal sito d’inizio, soprattutto nel caso di eucarioti multicellulari (**fig. 11.2D**). Negli eucarioti unicellulari quale il lievito, tali distanze sono più contenute, come si vede in **figura 11.2C**, che mostra la struttura di un tipico promotore di Pol II in lievito che contiene sequenze chiamate **UAS** (*Upstream Activating Sequence*) su cui si legano in modo specifico importanti proteine che influenzano l’efficienza di trascrizione. Descriveremo in dettaglio questi elementi e le relative proteine quando parleremo della regolazione della trascrizione negli eucarioti.

Fattori basali di Pol II e assemblaggio del complesso d'inizio

Vediamo ora come si assembla sul promotore il complesso d'inizio della trascrizione, promosso dai fattori basali. In **figura 11.6** è illustrata la regione prossimale di un promotore di Pol II con evidenziate le zone che legano i fattori basali. Questi comprendono più di 20 proteine che, in totale, hanno un peso molecolare superiore al milione di Dalton e si assemblano formando quello che viene chiamato **complesso di pre-inizio** o **PIC**. Il primo fattore a legarsi al promotore è TFIID (**fig. 11.6**), che ha dimensioni di circa 700 kDa e contiene la TBP (*TATA binding protein*) che abbiamo già incontrato precedentemente, insieme a numerosi, almeno 13, TAF (*TBP Associated Factors*). *In vitro*, TBP è sufficiente per promuovere una trascrizione basale a partire dal promotore; i fattori TAF sono importanti perché, come vedremo, sono il bersaglio di proteine attivatrici che aumentano il livello della trascrizione. Si noti che i TAF associati alla TBP sono diversi per i tre sistemi di trascrizione Pol I, Pol II e Pol III.

TBP è una proteina molto particolare che si lega come una “sella” sul DNA, interagendo con il solco minore. La molecola è un monomero con un asse di simmetria che la divide in due regioni simmetriche e, attraverso due loop molto conservati contenenti critici residui di fenilalanina, si lega alla TATA box, piegando il DNA di circa 80° come illustrato in **figura 11.7**. Questa piegatura facilita l'interazione di altri fattori nella regione del promotore; inoltre, recenti strutture cristallografiche dimostrano che tale curvatura permette al DNA del promotore di avvolgersi intorno alla polimerasi. Dopo TFIID si lega al complesso TFIIA

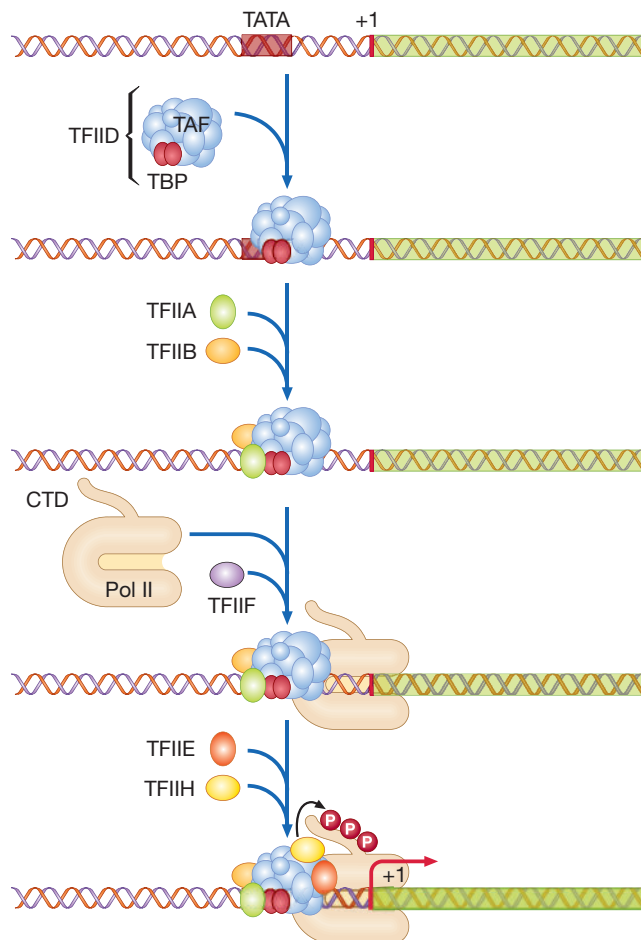
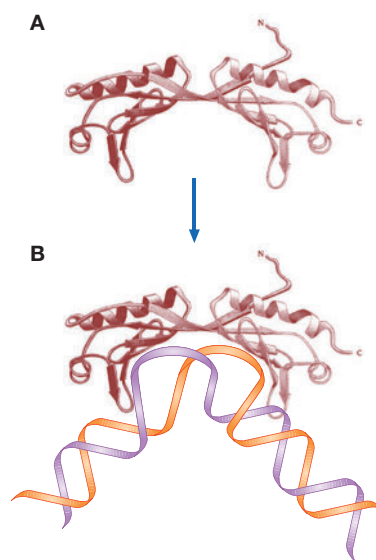


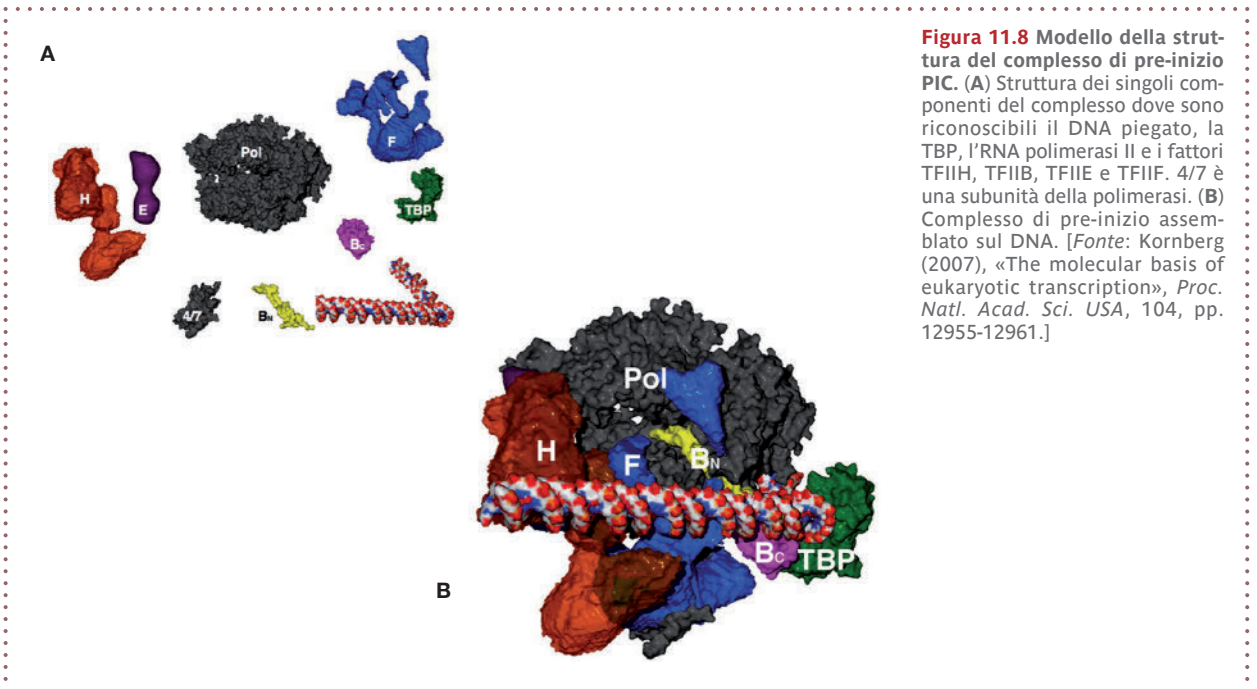
Figura 11.6 Formazione del complesso di pre-inizio (PIC) della RNA polimerasi II. La sequenza di eventi che porta alla formazione del complesso aperto per la Pol II inizia con il legame di TFIID, che è costituito dalla TBP e da numerose TAF. Successivamente, si legano al complesso TFIIA e TFIIB. A questo punto la polimerasi viene reclutata sul promotore insieme a TFIIF. Si legano poi TFIIIE e TFIIH, che con la sua attività elicastica apre il DNA e forma il complesso aperto. TFIIH è anche responsabile della fosforilazione della CTD di Pol II, che può, così, staccarsi dal promotore e proseguire nella fase di allungamento della trascrizione.

Figura 11.7 Struttura della TBP e sua interazione con la TATA box del promotore della RNA polimerasi II. (A) Struttura cristallografica della TBP, che è un monomero costituito da due domini simmetrici. (B) La TBP si lega come una sella al DNA in corrispondenza della TATA box, interagendo nel solco minore della doppia elica che viene allargato. Questo legame provoca una piegatura del DNA di circa 80°. TFIIB nel legarsi interagisce con le regioni a monte e a valle della TATA box e indirizza la polimerasi al sito di inizio della trascrizione.



(fig. 11.6), che ha la doppia funzione di stabilizzare il complesso TFIID-DNA e di impedire il legame di repressori che potrebbero interferire con la formazione del complesso di pre-inizio PIC. Successivamente, al complesso si lega TFIIB, che ha il compito di posizionare precisamente Pol II sul sito d'inizio, "misurandone" la distanza dalla TATA box. Infatti, mutazioni nel gene codificante TFIIB fanno sì che il trascritto abbia un sito d'inizio alterato. TFIIB interagisce con la TBP e il DNA, grazie alla piegatura indotta da TBP a monte e a valle della TATA box, è in grado d'interagire anche con la polimerasi, che si associa al complesso aiutata da TFIIF. TFIIB sembra avere un ruolo importante nelle fasi iniziali della sintesi dell'RNA: infatti, la risoluzione della struttura tridimensionale del complesso TFIIB-RNA polimerasi ha rivelato la presenza nella regione N-terminale di TFIIB di un dominio che, attraverso un lungo loop, si inoltra all'interno dell'enzima fino alla regione del suo sito attivo. Questo loop, denominato **dito B** (*B-finger*), si trova nella stessa regione dell'ibrido RNA-DNA nel complesso che sta trascrivendo e tale posizione non interferisce, ma anzi stabilizza la struttura dell'ibrido RNA-DNA. Quando l'RNA supera i 5 o 6 nt, deve competere con TFIIB per lo spazio nella cavità di Pol II. Se TFIIB ha la meglio, si ha una sintesi abortiva e la trascrizione deve re-iniziare; se, invece, la catena crescente di RNA riesce a superare i 6 nt, TFIIB viene espulso dalla cavità di Pol II, che si può così staccare dal promotore dando inizio alla fase di allungamento. Questi risultati hanno chiarito alcuni aspetti della fase abortiva della trascrizione: il B-finger ha un ruolo simile alla regione linker 3-4 del fattore σ nella polimerasi batterica. Anche il fattore TFIIF ha un ruolo simile al fattore σ : infatti, si lega a Pol II quando l'enzima è ancora libero da interazioni con altri fattori e ne impedisce il legame a regioni di DNA diverse dal promotore. Inoltre, TFIIF stabilizza le interazioni della polimerasi con TBP e TFIIB e sembra avere un ruolo nell'inizio della trascrizione, come indicato da recenti strutture cristallografiche che ne evidenziano una interazione con la bolla di trascrizione nel complesso aperto.

Come si vede in **figura 11.6**, la coda C-terminale (CTD) di Pol II, di cui abbiamo già parlato, è libera e non modificata. Quando, però, si forma il complesso con l'RNA polimerasi, intervengono altri due fattori, TFIIE e TFIIH. TFIIE ha la funzione di reclutare TFIIH nel complesso e TFIIH, che è costituito da varie subunità, ha un duplice ruolo. Il primo è quello di aprire il DNA



mediante la sua attività elicastica ATP-dipendente e di formare, quindi, il complesso di trascrizione “aperto”. L'altra funzione essenziale di TFIIFH è quella di fosforilare il CTD. Questa modificazione della coda CTD, cambiando la qualità delle interazioni tra il CTD e i vari componenti del PIC, promuove il distacco di Pol II dal promotore e permette l'inizio della fase di allungamento della trascrizione. Le modificazioni post-traduzionali della coda CTD sono anche essenziali per reclutare, nella fase di allungamento, le varie componenti che contribuiscono alla maturazione dell'RNA, come gli enzimi che formano il cappuccio (cap) e l'apparato di splicing (ne parleremo più avanti nei paragrafi 12.5 e 13.1). TFIIFH è il più grande tra i fattori basali di trascrizione essendo costituito da 10 subunità con un peso molecolare simile a quello dell'oloenzima Pol II. Il complesso TFIIFH gioca un ruolo importante anche in un meccanismo di riparazione del DNA, il NER (*vedi* par. 7.2) e, infatti, mutazioni in geni che codificano per alcune delle sue subunità sono collegate all'insorgenza di patologie, come *Xeroderma pigmentosum*, dovute a difetti nel NER. Questi dati, ovviamente, suggeriscono un legame fra trascrizione e riparazione del DNA. La struttura tridimensionale dei vari fattori del PIC e la loro associazione alla RNA polimerasi II sono mostrate in **figura 11.8**. La **tabella 11.1** riassume i fattori generali di trascrizione e le relative funzioni.

Ruolo del “mediatore” nella trascrizione di Pol II

I fattori basali e la polimerasi sono in grado di trascrivere *in vitro* un DNA nudo a partire dal promotore. Nelle cellule *in vivo*, però, occorrono altri componenti che, da una parte, modificano la cromatina per permettere al DNA di essere trascritto e, dall'altra, portano a livelli di trascrizione elevati. Uno dei componenti più importanti per la formazione del PIC *in vivo* è il complesso chiamato **mediatore** (**finestra 11.1**), che è costituito da circa 20 subunità, delle quali più di 7 sono conservate dal lievito all'uomo. Il mediatore, la cui struttura è stata risolta a bassa risoluzione mediante tecniche di microscopia elettronica (**fig. 11.9A**), si lega al complesso d'inizio trasportandovi, come mostrato in **figura 11.9B**, l'RNA polimerasi. Questa è legata direttamente al mediatore che interagisce con

Tabella 11.1 Fattori generali di trascrizione della RNA polimerasi II.

Fattore	Numero di subunità	Peso molecolare delle subunità in kDa	Funzioni
TFIIA	3	12, 19, 35	Stabilizza il legame di TBP e TFIIIB
TFIIB	1	25	Lega TBP, seleziona il sito di inizio della trascrizione e recluta Pol II
TFIID	12	Da 15 a 250	Interagisce con i fattori di regolazione ed è formata da TBP e TAF
TBP	1	38	Subunità di TFIID, si lega in maniera specifica alla TATA box
TFIIE	2	34, 57	Recluta TFIIH nel complesso di pre-inizio
TFIIF	2	30, 74	Lega Pol II e TFIIB
TFIIH	9	Da 38 a 98	Ha un'attività elicasica che apre il promotore, fosforila la coda C-terminale (CTD) di Pol II
Pol II	12	Da 10 a 220	Catalizza la sintesi dell'RNA usando il DNA come stampo
Totale	42	Più di 1000	

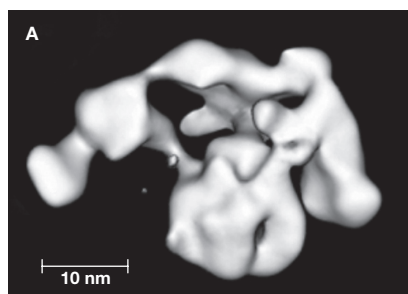
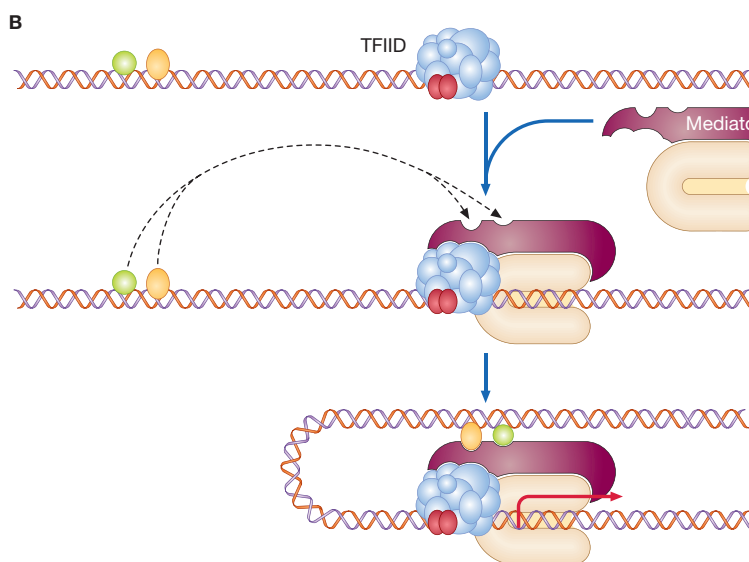


Figura 11.9 Struttura della RNA polimerasi II in complesso con il mediatore (oloenzima). (A) Struttura del complesso RNA polimerasi/mediatore, determinata tramite microscopia elettronica. (B) Il complesso del mediatore aiuta la polimerasi a legarsi ai promotori attraverso l'interazione con fattori già legati agli elementi vicini al sito d'inizio (ad es. TFIID) e agli enhancer (indicati dai simboli in verde e giallo). L'interazione con TFIID è essenziale per assemblare un complesso attivo; si pensa che TFIID rimanga legato al promotore per facilitare il successivo assemblaggio di un altro complesso di trascrizione. [Fonte: Chadick, J.Z. e Asturias, F.J. (2005), *TRENDS in Bioch. Sci.*, 5.]



FINESTRA 11.1 IL “MEDIATORE” E LA SUA SCOPERTA

Dopo la scoperta degli attivatori trascrizionali e della loro struttura con domini di legame al DNA e domini di attivazione, si è subito cercato di capire chi fosse il bersaglio di questi fattori, oltre all'RNA polimerasi II stessa. All'inizio si è pensato che i fattori basali di trascrizione e gli attivatori interagissero direttamente e che la loro interazione fosse sufficiente non solo a iniziare accuratamente la trascrizione, ma anche a regolarla. Però, in quegli anni, nel laboratorio di Mark Ptashne, sono state ottenute le prime evidenze dell'esistenza e importanza di un altro componente, chiamato **mediatore**. Infatti, studiando gli effetti dell'attivatore Gal4 sulla trascrizione, i ricercatori avevano notato che la sovrapproduzione in lievito dell'attivatore Gal4 interferiva con l'attivazione di Pol II da parte di un altro attivatore. Questo effetto, definito **squelching** (sequestrare), fu interpretato come una competizione tra i due attivatori per un bersaglio comune presente nelle cellule in quantità limitante. Si è inizialmente pensato che il bersaglio potesse essere un componente dell'apparato trascrizionale, cioè la polimerasi stessa o uno dei fattori basali di trascrizione. Il laboratorio di Roger Kornberg (premio Nobel per i suoi studi sulla trascrizione utilizzando componenti purificati *in vitro*) scoprì però che un altro componente fino allora ignoto era necessario per un'accurata trascrizione *in vitro*.

Questi esperimenti, inoltre, misero in dubbio l'ipotesi che gli attivatori trascrizionali contattino direttamente l'apparato basale di trascrizione. Infatti, l'aggiunta di grandi quantità di Pol II o dei fattori basali di trascrizione (GTF) non riusciva a eliminare lo squelching e a far ripartire l'attivazione, mentre ciò si otteneva con l'aggiunta di una frazione proteica che non conteneva né GTF né Pol II.

L'attività di questa frazione è stata definita, da allora, “mediatore”. Dopo molti tentativi, tutti i componenti dell'apparato trascrizionale sono stati isolati e purificati e il complesso attivo del mediatore è risultato essere composto da almeno 20 subunità, con una massa totale di circa un milione di Dalton. Tredici subunità sono costituite dai prodotti di geni già identificati e coinvolti nella regolazione della trascrizione. Successivamente, il mediatore è stato isolato anche da estratti di cellule di mammifero e si è visto che questo complesso è estremamente conservato.

Dopo la sua purificazione si è riusciti a stabilire che TFIID e i TAF servono al riconoscimento del promotore, ma non alla regolazione della trascrizione. Dai risultati più recenti ottenuti sul mediatore sono chiari tre punti: (1) il mediatore è assolutamente richiesto per la trascrizione da praticamente tutti i promotori di Pol II e questo complesso è necessario quanto la polimerasi stessa; (2) il mediatore interagisce direttamente con gli attivatori e la polimerasi, formando complessi stabili con gli attivatori legati all'enhancer e, successivamente, contattando Pol II e i fattori basali per stimolare l'inizio della trascrizione; (3) il mediatore è anche coinvolto nella repressione della trascrizione. In pratica, in tutti gli eucarioti il mediatore è richiesto per trasdurre dall'enhancer al

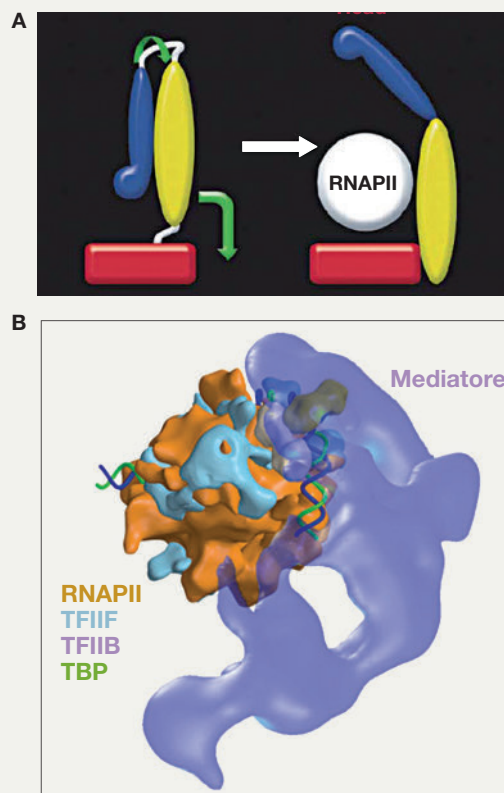


Figura 11F1.1 La struttura del mediatore è stata determinata con studi di microscopia elettronica. (A) Modello schematico della struttura del mediatore in cui si vedono, in diversi colori, i vari domini chiamati testa, parte intermedia e coda. La coda va incontro a grandi cambiamenti conformazionali quando è legata alla polimerasi. (B) Modello derivato dalla ricostruzione delle immagini al microscopio elettronico in cui si individuano la polimerasi e alcuni fattori di trascrizione. [Fonte: Chadick, J.Z. e Asturias, F.J. (2005), *TRENDS in Biochem. Sci.*, 5.]

promotore i segnali regolativi degli attivatori e dei repressori. Recentemente, con tecniche di microscopia elettronica ad alta risoluzione, è stato possibile vedere il mediatore in azione. In **figura 11F1.1A** si vede il modello di due possibili configurazioni del mediatore. Prima di interagire con la polimerasi e i fattori GTF, il mediatore ha una conformazione chiusa; successivamente, quando contatta Pol II e i GTF, ha una conformazione aperta. Non è ancora chiaro come questi due stati conformazionali possano essere indotti, ma un'ipotesi possibile è che il contatto con gli attivatori induca il cambio conformazionale, permettendo al mediatore di legarsi alla polimerasi. In un secondo modello, la polimerasi stessa può indurre la modificazione per formare il complesso insieme ai GTF. In **figura 11F1.1B** si vede il complesso del mediatore e alcuni GTF, dopo la ricostruzione delle immagini, acquisite al microscopio elettronico.

il CTD, mascherandolo. Il principale ruolo del mediatore è quello di fare da ponte tra gli attivatori legati a monte e il complesso d'inizio e, inoltre, alcune sue subunità possono avere un ruolo nella repressione della trascrizione. Il mediatore stimola la trascrizione in un sistema purificato, anche in assenza di un attivatore; questa stimolazione può derivare dalle strette interazioni che ha con la polimerasi, i fattori basali e il DNA. Questo stretto contatto sembra stabilizzare il complesso di pre-inizio, promuovendo la sua formazione e il suo mantenimento per più cicli di trascrizione. Poiché il mediatore non è strettamente necessario per la trascrizione basale *in vitro*, si è a lungo discusso se lo si dovesse considerare come un fattore basale di trascrizione o come un co-attivatore. Generalmente il mediatore è considerato come un componente essenziale del complesso di pre-inizio, dato che alcune delle sue subunità sono necessarie per la trascrizione di quasi tutti i geni trascritti da Pol II. Il mediatore è un complesso proteico presente esclusivamente negli eucarioti e alcune sue funzioni nella regolazione della trascrizione sono ancora da chiarire.

Struttura di Pol II e reazione di allungamento nella sintesi dell'RNA

Dopo aver discusso la formazione del complesso di pre-inizio (PIC), descriviamo la struttura dell'RNA polimerasi II per cercare di capire come avviene a livello molecolare la sintesi dell'RNA da parte di tale enzima. Come abbiamo già descritto nel paragrafo 9.3 per l'RNA polimerasi procariotica, anche Pol II è costituita da più subunità che danno a questo enzima una forma a "pinza" che si apre e chiude sul DNA inglobandolo in una lunga scanalatura nella struttura dell'oloenzima (*vedi* fig. 9.8). La parte superiore della "pinza" è quella più mobile e permette l'ingresso del DNA. La brusca curvatura che, a causa della struttura del "muro", il filamento stampo del DNA è costretto a fare, ha come risultato quello di esporre la base non appaiata nel sito attivo, pronta ad appaiarsi con il nucleotide successivo che entrerà nel complesso. In **figura 11.10A** è mostrato, a sinistra, uno spaccato della Pol II e, a destra, un ingrandimento del sito attivo, costituito dal sito E (entrata), da cui entrano i nucleotidi, e dal sito A (assemblaggio). In **figura 11.10B** si può vedere meglio un nucleotide (giallo) nel sito A pronto per appaiarsi al filamento stampo e un nucleotide (blu) nel sito E, che deve ancora ruotare per poter esporre la base per l'appaiamento. In verde si vede l' α -elica "ponte" che connette le due grandi subunità dell'enzima.

Ma come fa l'RNA polimerasi a riconoscere il giusto nucleotide che deve essere aggiunto alla catena di RNA? Questo aspetto è cruciale per la fedeltà di lettura del codice genetico. Recentemente, la risoluzione di una struttura cristallografica dell'RNA polimerasi II di lievito contenente 1 NTP ha evidenziato la presenza di un dominio proteico chiamato **trigger loop** (grilletto). Questo ha il ruolo di riconoscere e discriminare il nucleotide corretto per evitare errori nella sintesi; inoltre, promuove la formazione del legame fosfodiesterico e l'avanzamento della polimerasi (**figura 11.11**). Il trigger loop posiziona il nuovo NTP nel sito attivo in modo tale che, se l'appaiamento è corretto, il nucleotide riesce a posizionarsi correttamente rispetto al 3'OH della molecola di RNA nascente e al nucleotide complementare nel filamento stampo del DNA. Il trigger loop promuove l'attacco nucleofilo da parte del 3'OH e il distacco del pirofosfato. Inoltre, ha anche il ruolo di promuovere l'avanzamento dell'enzima: infatti, dopo che si è formato il legame fosfodiesterico, questa struttura ritorna nella posizione iniziale, proprio come un grilletto, riformando il sito A vuoto e interagendo con altri domini proteici come il bridge, che connettendo le due subunità grandi della pinza promuove l'avanzamento dell'enzima. Va notato come, nonostante non ci siano delle omologie strutturali tra DNA e RNA polimerasi, ci siano forti somiglianze nel meccanismo, come la presenza degli ioni Mg^{2+} che trattengono i gruppi fosfato e la necessità delle interazioni di impilamento con la base precedentemente

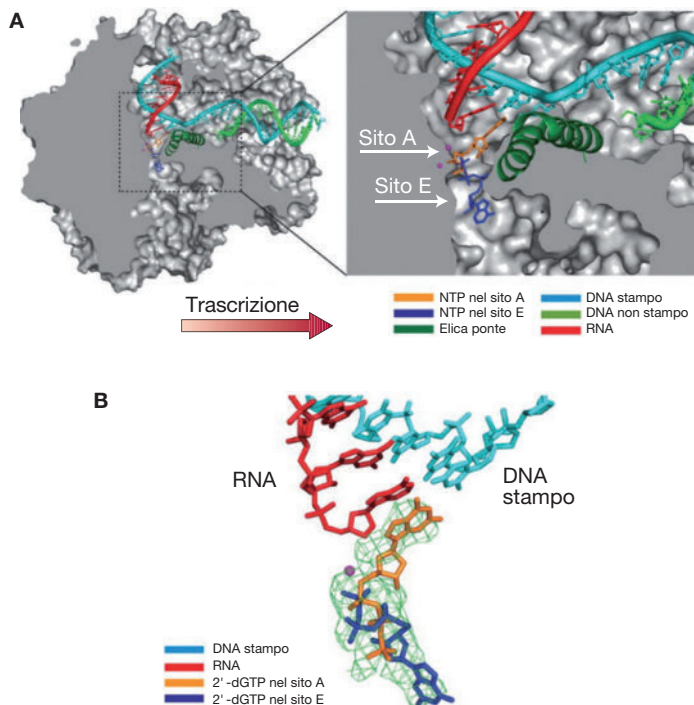
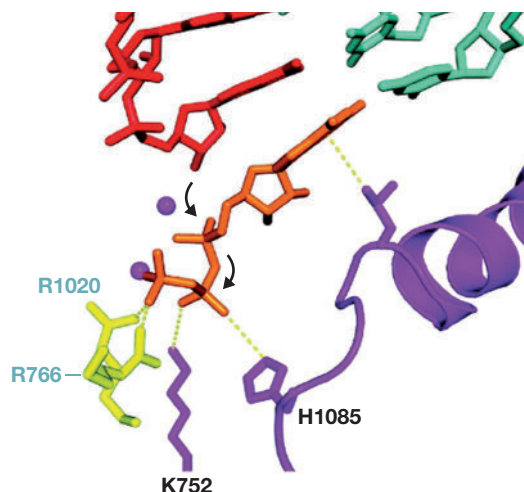


Figura 11.10 Struttura del sito attivo dell'RNA polimerasi II. (A) A sinistra, spaccato della RNA polimerasi II; a destra ingrandimento del sito attivo. (B) Dettaglio del posizionamento dei nucleotidi nel sito attivo: dal codice colori è possibile riconoscere i nucleotidi, il DNA stampo, l'RNA e l'elica ponte (bridge helix) che connette le due grandi subunità della polimerasi. [Fonte: Kornberg L. (2007), «The molecular basis of eukaryotic transcription», *Proc. Natl. Acad. Sci. USA*, 104, pp. 12955-12961.]

aggiunta al polinucleotide nascente, per il corretto riconoscimento del nucleotide da aggiungere e la formazione del legame fosfodiesterico.

Una volta che la polimerasi ha superato la sintesi abortiva e si è staccata dal promotore e dalla maggior parte dei fattori basali, inizia la fase di allungamento della trascrizione. In questa fase, il CTD fosforilato recluta alcuni fattori specifici che, come menzionato sopra, sono necessari per il processamento, lo splicing e la poliadenilazione dell'RNA. Da un punto di vista strutturale, la coda CTD si trova vicino al canale di uscita dell'RNA e la sua forma non strutturata la fa estendere oltre l'enzima, a seconda del numero di ripetizioni, fino a circa 700-800 Å. Questa caratteristica permette alla coda di reclutare i componenti sopra descritti e di dirigerli sull'RNA che sta uscendo dall'enzima.

Figura 11.11 Trigger loop e suo ruolo nella trascrizione. Il trigger loop (viola), mediante le interazioni di alcuni residui laterali come l'istidina 1085, posiziona l'NTP (arancio) esattamente nel sito attivo. Se l'appaiamento è corretto, il nucleotide riesce a formare interazioni d'impilamento con la catena dell'RNA (rosso), ad avvicinarsi alla giusta distanza al 3'OH dell'RNA e di posizionarsi correttamente rispetto al filamento stampo del DNA (verde). Inoltre, l'istidina promuove un flusso di elettroni che permettono l'attacco nucleofilo da parte del 3'OH e il distacco del pirofosfato, indicati con delle frecce nere. [Fonte: Kornberg, L. (2007), «The molecular basis of eukaryotic transcription», *Proc. Natl. Acad. Sci. USA*, 104, pp. 12955-12961.]



Analogamente alla polimerasi batterica, durante la fase di allungamento Pol II va incontro a delle pause che dipendono anche dalla sequenza che viene trascritta. Come nei procarioti, esistono dei fattori di allungamento che permettono alla polimerasi di procedere e superare questi eventuali rallentamenti sul DNA. Oltre a TFIIF, che sembra avere un ruolo nello stimolare l'allungamento iniziale, il fattore TFIIS limita il tempo delle pause che la polimerasi è costretta a fare sul DNA stampo. TFIIS funziona in modo simile a GreB, che abbiamo descritto nel paragrafo 9.4 come fattore di allungamento della polimerasi batterica. TFIIS ha una lunga struttura che si inserisce, attraverso l'"imbuto" da dove entrano i nucleotidi, fino al sito attivo, dove avviene la sintesi dell'RNA. TFIIS ha anche la funzione di stimolare l'attività di correzione da parte della polimerasi nel caso sia stato incorporato un nucleotide sbagliato. Queste reazioni di correzione possono, come nei batteri, essere realizzate attraverso la reazione inversa di sintesi dell'ultima base, con la polimerasi che indietreggia sullo stampo o con una degradazione limitata della catena di RNA. TFIIS stimola l'attività ribonucleasica della polimerasi, che rimuove le basi sbagliate e poi riprende a trascrivere dopo la pausa. È interessante notare che TFIIS, nonostante abbia sequenza e struttura diverse da GreB, interagisce con la polimerasi in modo molto simile.

11.5 Regolazione

Nei paragrafi precedenti abbiamo descritto i meccanismi attraverso i quali le tre RNA polimerasi eucariotiche riconoscono i loro promotori e trascrivono diversi tipi di RNA. Analizziamo ora i meccanismi che regolano la trascrizione della RNA polimerasi II. Tale regolazione si realizza attraverso l'integrazione dei ruoli di numerose proteine, genericamente chiamate **fattori di regolazione**, che, controllando sia positivamente che negativamente la trascrizione di geni specifici, sovrintendono al normale funzionamento della cellula, al differenziamento dei vari tessuti, allo sviluppo di un organismo. La funzione di molti regolatori è, a sua volta, influenzata da stimoli esterni, permettendo così all'organismo di reagire a specifici segnali.

Questo complesso apparato di attivatori, co-attivatori e repressori permette il controllo della trascrizione della maggior parte degli mRNA, sia stimolando e aumentando il livello basale di trascrizione, che modificando lo stato della cromatina, così che le regioni di controllo di specifici geni diventino accessibili all'apparato trascrizionale solo in determinati momenti della vita della cellula e dell'organismo. Vedremo, inoltre, che questi meccanismi di controllo, pur nelle prevedibili differenze, sono conservati negli eucarioti, dal lievito all'uomo.

Struttura modulare dei promotori eucariotici

Come abbiamo visto in **figura 11.2**, un promotore di Pol II è costituito da vari elementi che, schematizzando, si possono suddividere in due gruppi distinti: quelli prossimali al sito d'inizio della trascrizione (fino a circa 200 pb), a cui si legano i fattori basali di trascrizione, la polimerasi e alcuni attivatori, e quelli distali, che si possono trovare anche a 100 kb di distanza, ai quali si legano i fattori di regolazione della trascrizione.

Anche nella regione prossimale del promotore vi sono elementi sul DNA che legano attivatori o repressori. Nel lievito la maggior parte dei geni ha un elemento chiamato **UAS** (*Upstream Activating Sequence*), che si trova da -100 a -200 pb dal sito d'inizio della trascrizione, al quale si legano attivatori specifici per il gene posto a valle (**fig. 11.2C**). Negli eucarioti multicellulari il promotore è ricco di sequenze regolatrici costituite da parecchi elementi prossimali posti fino a -200 pb dal sito d'inizio (**fig. 11.2D**). Queste sequenze di DNA sono state individuate attraverso esperimenti di mutagenesi sito-specifica nella regione del promotore, spesso usando un gene reporter per saggiare l'attività trascrizionale,

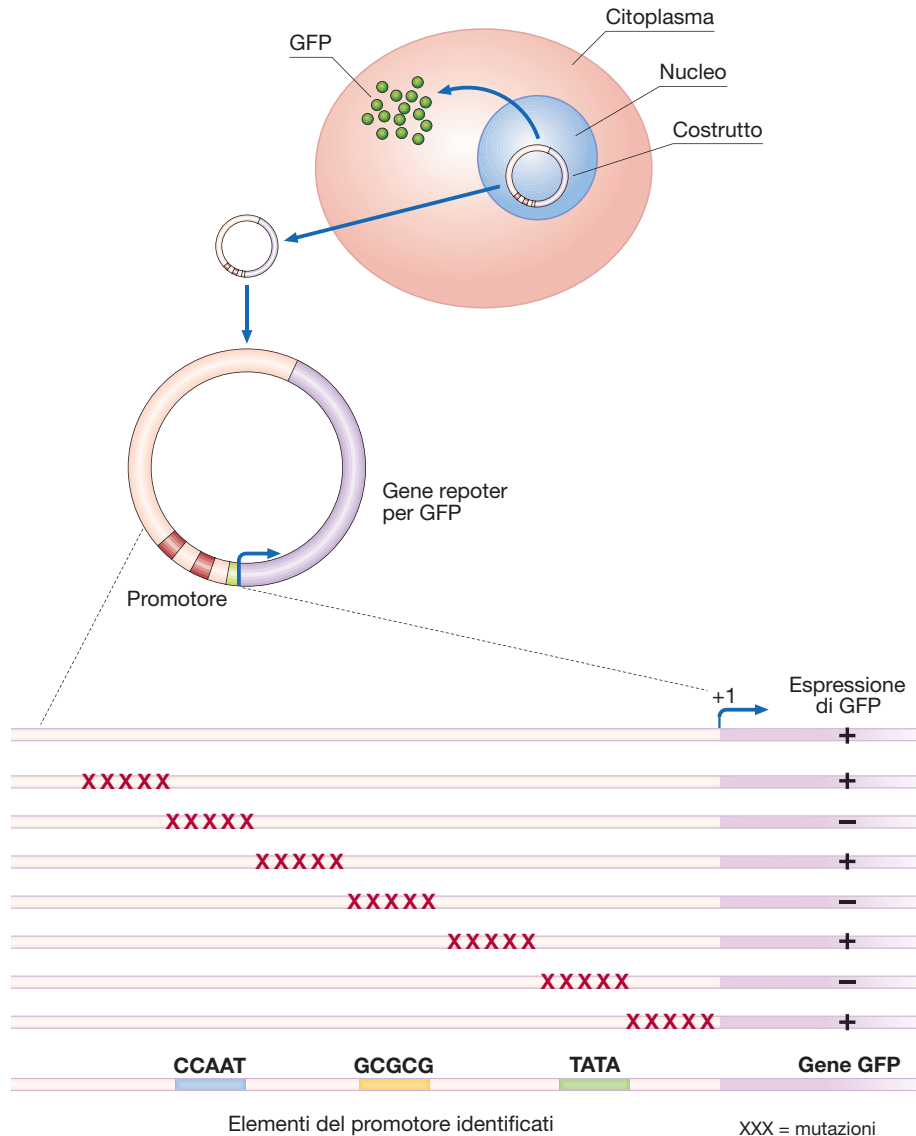


Figura 11.12 Identificazione delle sequenze rilevanti in un promotore trascritto da Pol II. Le sequenze più importanti di un promotore vengono identificate con esperimenti come quello qui brevemente descritto (e discusso più approfonditamente nel par. 20.5). Il promotore da analizzare viene posizionato a monte di un gene "reporter", per esempio GFP che codifica per una proteina fluorescente (GFP) la cui presenza può essere facilmente misurata

con un semplice saggio. Il costrutto chimerico viene introdotto per trasfezione in cellule in coltura e la sua espressione misurata con un saggio che rivela la presenza della proteina fluorescente. Una serie di mutazioni, indicate in figura con le X, vengono saggate per il loro effetto sulla produzione di GFP. Mutazioni che riducono l'espressione del gene reporter individuano elementi regolativi importanti.

come illustrato nel capitolo 20 (vedi par. 21.10 e fig. 20.53) e, più sinteticamente, in **figura 11.12**. In questo caso, un promotore eucariotico viene posizionato a monte del gene che codifica per la GFP (*Green Fluorescent Protein*, proteina fluorescente verde) e il plasmide che contiene questo costrutto è trasfettato in cellule eucariotiche in coltura. Una serie di mutazioni che coprono la regione prossimale permette di individuare le sequenze del promotore su cui si legano i fattori basali e di regolazione essenziali per la sua trascrizione. Come esempio, in **figura 11.13** è mostrato il promotore del gene codificante la metallotioneina, con i relativi elementi di controllo così identificati.

I siti di riconoscimento prossimali nel promotore sono, in generale, formati da sequenze nucleotidiche che si presentano a blocchi (box), quali, per esempio,

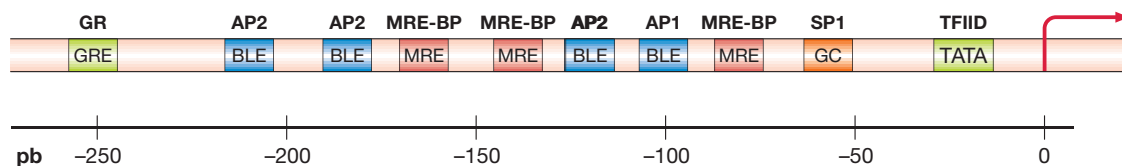


Figura 11.13 Struttura del promotore del gene per la metallotioneina IA umana. Il promotore del gene per la metallotioneina è riportato come esempio di organizzazione modulare dei promotori eucariotici. L'attività basale di questo promotore (cioè in assenza di metalli) è dovuta alla TATA box e al complesso di trascrizione ad essa legato: un fattore SP1 legato alla GC-box (GGGCGG) e i

fattori AP1 e AP2 legati a delle box BLE (*Basal Level Enhancer*). La trascrizione indotta da metallo richiede alcune box MRE (*Metal Responsive Element*), la cui attività dipende dal fattore MRE-BP (*MRE Binding Protein*). L'attivazione da parte dei glucocorticoidi dipende dal legame del recettore dei glucocorticoidi (GR) a un elemento di risposta ai glucocorticoidi (GRE).

la CAAT box e la GC box. Alla CAAT box si lega il fattore trascrizionale noto come CBF (*CAAT Binding Factor*) o NF-Y; la CAAT box può funzionare in entrambi gli orientamenti e si ritiene che influenzi l'efficienza di un promotore, ma non abbia un ruolo nel determinarne la specificità trascrizionale. Alla GC box (GGGCGG) si lega il fattore Sp1, presente in molti organismi compreso l'uomo, che è stato originariamente identificato come necessario per la trascrizione dei geni precoci e tardivi del virus SV-40, ma che è coinvolto nell'espressione di molti geni necessari per la vitalità delle cellule definiti **housekeeping**. Molti elementi prossimali sono presenti in geni che vengono trascritti costitutivamente, ma anche nei promotori di geni coinvolti nella risposta a vari tipi di stress o all'esposizione a metalli pesanti come, per esempio, il gene della metallotioneina, mostrato in **figura 11.13**, che contiene siti MRE (*Metal Response Element*), GRE (*Glucocorticoid Response Element*) e siti di legame per i fattori trascrizionali Sp1, Ap1 e Ap2. Un'altra frequente sequenza di risposta è l'ottamero (Oct), costituito, appunto, da 8 pb. A questo elemento si lega, in genere, la proteina Oct-1 che è un attivatore che agisce su molti promotori, come quelli per l'attivazione del gene per l'istone H2B. Nelle cellule della linea linfocitaria, invece, al posto di Oct-1 si lega all'ottamero la proteina Oct-2, che attiva specificamente il gene per la catena leggera delle immunoglobuline. Questo dato indica che Oct-2 è un attivatore specifico, mentre Oct-1 può attivare geni diversi. Entrambi, tuttavia, si legano allo stesso elemento in *cis* sul DNA; la loro specificità d'azione non è, quindi, legata solo a dove si legano sul promotore, ma piuttosto al numero e al tipo d'interazioni che sono in grado d'instaurare con altri fattori trascrizionali.

Gli elementi distali che si chiamano **enhancer** (amplificatori) contengono molti elementi o sequenze di riconoscimento alle quali si legano i fattori di trascrizione, per lo più con funzione attivatrice. In **figura 11.14** si può osservare la formazione di un complesso di attivatori che si legano a un enhancer per formare un complesso denominato **enhanceosoma**. Questo complesso proteico entra in comunicazione con le regioni prossimali del promotore, attraverso la formazione di un'ansa sul DNA, allorché gli attivatori legati all'enhancer interagiscono con il "mediatore", che, a sua volta, richiama la RNA polimerasi II e ne facilita il legame ai fattori basali di trascrizione. Va precisato, però, che molti dei meccanismi che influenzano questa regolazione a distanza sono ancora in gran parte da chiarire a livello molecolare. Alcuni dati suggeriscono che regioni apparentemente molto distanti a livello di sequenza di DNA, sono in realtà fisicamente vicine a causa della struttura ad anse dei cromosomi eucariotici, spiegando così almeno in parte l'azione di enhancer molto lontani. Questi elementi *cis*-agenti sono di solito posizionati a 700-1000 pb dal sito d'inizio della trascrizione, hanno una lunghezza intorno alle 500 pb e contengono in media 10-12 siti di legame per fattori di trascrizione diversi. Nel caso della regolazione dei geni per i recettori degli odori nel topo sembra addirittura esistere un enhancer che regola geni che si trovano su cromosomi diversi.

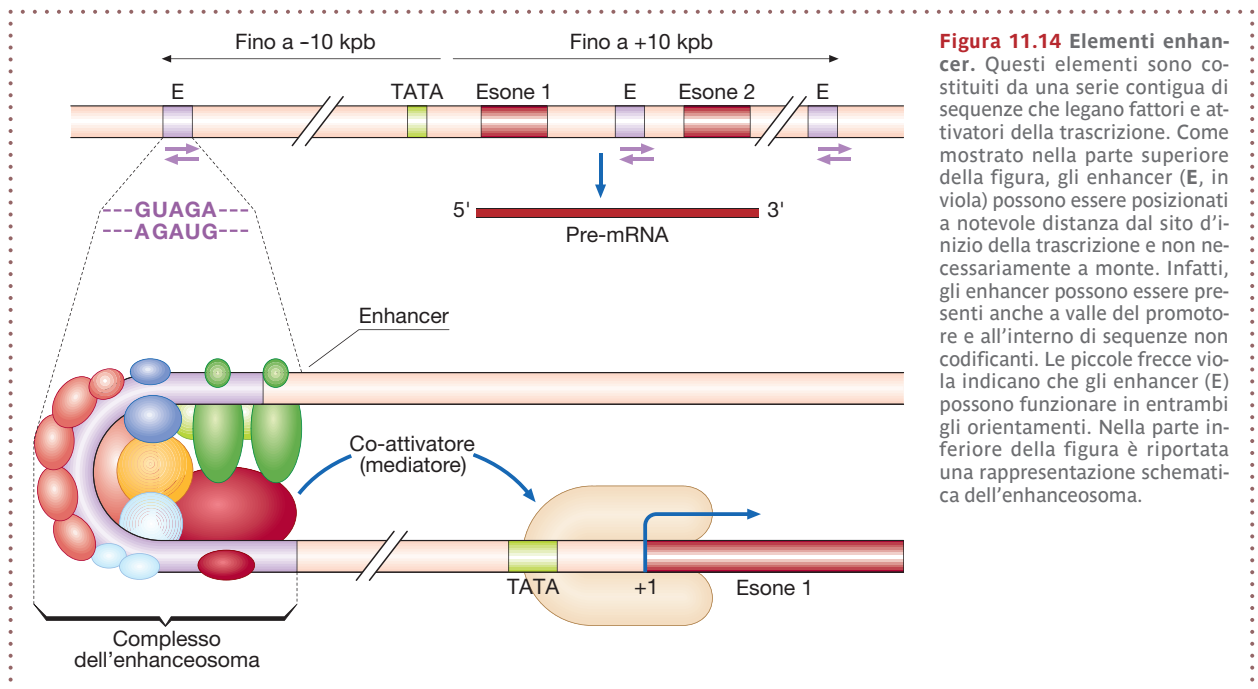


Figura 11.14 Elementi enhancer. Questi elementi sono costituiti da una serie contigua di sequenze che legano fattori e attivatori della trascrizione. Come mostrato nella parte superiore della figura, gli enhancer (E, in viola) possono essere posizionati a notevole distanza dal sito d'inizio della trascrizione e non necessariamente a monte. Infatti, gli enhancer possono essere presenti anche a valle del promotore e all'interno di sequenze non codificanti. Le piccole frecce viola indicano che gli enhancer (E) possono funzionare in entrambi gli orientamenti. Nella parte inferiore della figura è riportata una rappresentazione schematica dell'enhanceosoma.

Alcune caratteristiche degli enhancer sono illustrate in **figura 11.15**:

1. possono funzionare in entrambi gli orientamenti rispetto alla direzione della trascrizione (**fig. 11.15A e B**);
2. possono trovarsi sia a monte che a valle del gene che controllano;
3. possono essere localizzati negli introni dello stesso gene che viene controllato.

Gli enhancer contengono, oltre a sequenze specifiche per gli attivatori trascrizionali, anche sequenze che, solitamente, si trovano nella regione prossimale del promotore, come l'ottamero e i siti di legame per Ap1, Ap2 e Sp1.

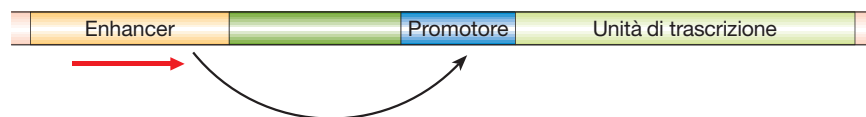
La specificità della trascrizione può essere controllata sia dalla regione del promotore che dall'enhancer. In alcuni casi, la specificità trascrizionale è data dalla regolazione a livello del promotore e l'enhancer è usato per aumentare il livello del messaggero prodotto; in altri casi, il promotore del gene è attivato solo se l'enhancer che lo controlla è a sua volta attivato. Esistono due teorie principali sul tipo di informazione contenuta negli enhancer. Secondo la prima teoria i fattori trascrizionali si legano ad essi in modo cooperativo, coordinato e specifico per formare l'enhanceosoma. A supporto di questo modello si è osservato che mutazioni puntiformi che alterano il legame di un singolo componente hanno un effetto evidente sull'azione globale dell'enhancer. Nella seconda teoria gli enhancer sono visti come *flexible billboards*, cioè come "tabelloni flessibili" dove i vari attivatori e co-attivatori si legano in modo indipendente gli uni dagli altri e il loro effetto combinato è "letto" dal "mediatore" e dall'apparato basale di trascrizione, quando la zona dell'enhancer entra in contatto con la zona prossimale del promotore.

Oltre agli enhancer esistono elementi definiti **silencer** che funzionano in modo simile, legando, invece che attivatori, repressori della trascrizione: come gli enhancer, possono trovarsi a valle o a monte dei geni che controllano. Questi repressori possono agire in modi diversi: per esempio, possono mascherare la regione di attivazione agli attivatori o competere con il legame sul mediatore.

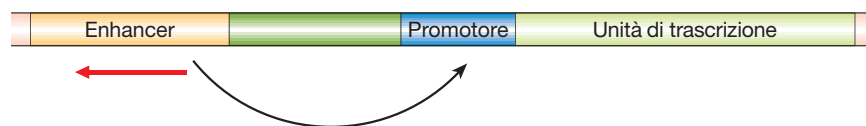
Un altro gruppo di elementi di controllo che agiscono in *cis* sul DNA è rappresentato dagli **isolatori**. Questi elementi (**fig. 11.16**) sono posizionati in modo tale da influenzare l'azione degli enhancer contigui. In alcuni casi, infatti, sono utilizzati per isolare in una direzione l'azione di un enhancer che potrebbe in-

Figura 11.15 Gli enhancer possono funzionare in diversi modi. Gli enhancer possono funzionare (A) a distanza, (B) in ambedue gli orientamenti e (C) in diverse posizioni, anche all'interno dell'unità di trascrizione.

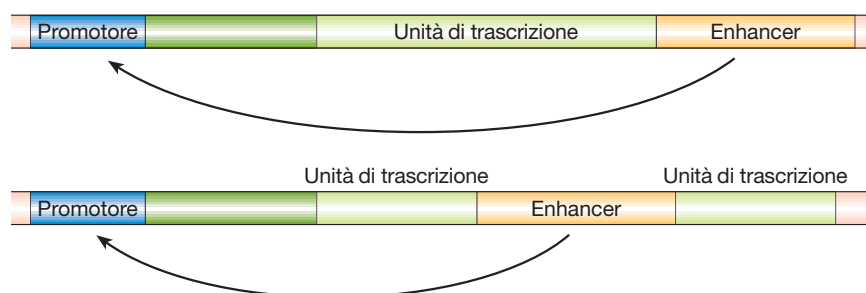
A Distanza



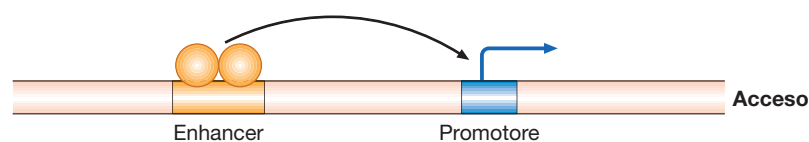
B Orientamento



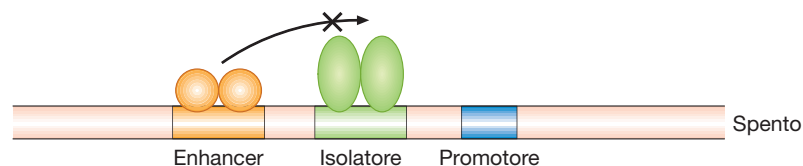
C Posizione



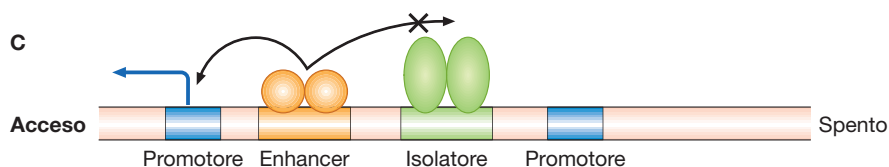
A



B



C



D

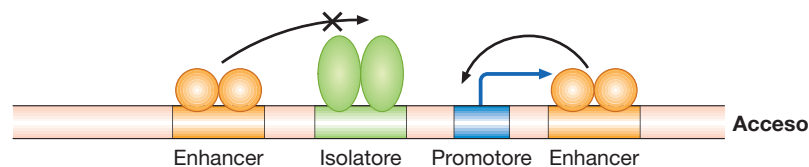


Figura 11.16 Ruolo degli isolatori. Gli isolatori sono sequenze di DNA a cui si legano specifiche proteine che influenzano l'azione degli enhancer. (A) Un enhancer attiva il promotore posto a valle. (B) Un isolatore impedisce la comunicazione dell'enhancer con il promotore. (C) Anche in presenza di un isolatore che impedisce

la comunicazione con un promotore a valle, l'enhancer è in grado di attivare un promotore posto a monte della sua posizione. (D) L'isolatore impedisce l'attività di un enhancer posto a monte del promotore, ma non reprime l'attività di un enhancer che si trova a valle del promotore.

fluenzare due promotori diversi. Gli isolatori hanno una doppia funzione: (1) agiscono impedendo l'attivazione o la repressione di un gene da parte di un enhancer o di un silencer; (2) agiscono come “marcatori” di confine tra una zona di eterocromatina e una zona di eucromatina. Un isolatore contiene siti di legame per proteine che, nel caso della struttura della cromatina, impediscono il propagarsi e l'espandersi di regioni eterocromatiche.

Il quadro generale della distribuzione degli elementi in *cis* sul DNA e del legame a tali sequenze di regolatori trascrizionali sia nelle regioni prossimali che distali dei promotori, mette in evidenza la natura combinatoria della regolazione della trascrizione negli eucarioti. La funzione di regolatori tessuto-specifici o di risposta a stimoli si integra con quella di attivatori generali e fattori basali nel controllare con precisione e senza ambiguità la trascrizione di un determinato gene. Questo meccanismo combinatorio è il risultato del legame estremamente specifico sul DNA degli attivatori *trans*-agenti.

11.6 Struttura modulare e domini dei fattori di trascrizione (transattivatori)

Descrivendo il controllo del ciclo vitale del fago lambda (*vedi* par. 10.4) abbiamo visto che la proteina CI (repressore di lambda) attiva la trascrizione del suo stesso gene: infatti, la proteina CI legata a O_{R2} attiva la trascrizione, interagendo con il complesso dell'RNA polimerasi, e il sito di attivazione del repressore CI non è molto lontano dai domini con cui CI si lega al DNA. In modo analogo, gli attivatori della trascrizione degli eucarioti contengono un sito di legame al DNA e un sito di attivazione, ma nella maggior parte dei casi questi due domini appartengono a due regioni diverse della proteina.

Uno degli attivatori trascrizionali eucariotici più studiati, e che è servito come modello a tutti gli altri, è la proteina **Gal4** (**fig. 11.17**), che è richiesta per attivare la trascrizione dei geni per l'utilizzo dello zucchero galattosio nel lievito *S. cerevisiae*, legandosi, per esempio, al promotore del gene *GAL1*. Come mostrato in **figura 11.18**, Gal4 si lega a 4 siti che contengono ciascuno una sequenza di 17 nt e che fanno parte della regione UAS, situata a circa -270 pb a monte del sito d'inizio della trascrizione del gene *GAL1*. In presenza di Gal4 la trascrizione di tale gene aumenta di circa 1000 volte. Due esperimenti molto eleganti hanno chiarito il funzionamento di Gal4 e hanno definito la natura generale degli attivatori della trascrizione negli eucarioti. Usando come reporter il gene batterico per la β -galattosidasi (*lacZ*), è stato possibile analizzare gli effetti di diverse proteine artificiali che avevano come base la proteina Gal 4. Nello schema mostrato in **figura 11.19A**, Gal4 wild-type si lega al suo sito specifico e attiva la trascrizione del gene *lacZ*. Questo è facilmente visibile perché le cellule delle colonie di lievito in cui sono presenti i vettori che contengono il gene reporter e le varianti di Gal4 da studiare diventano di colore blu se il gene *lacZ* viene espresso. Se si elimina la regione che contiene il dominio di attivazione trascrizionale, Gal4 si lega al DNA, ma non è in grado di attivare la trascrizione. Con questo esperimento è stato individuato il dominio di attivazione trascrizionale di Gal4.

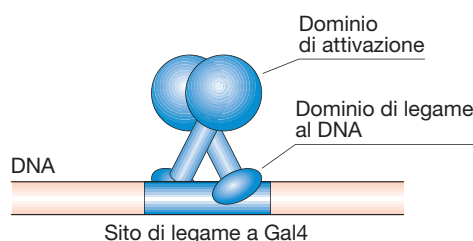
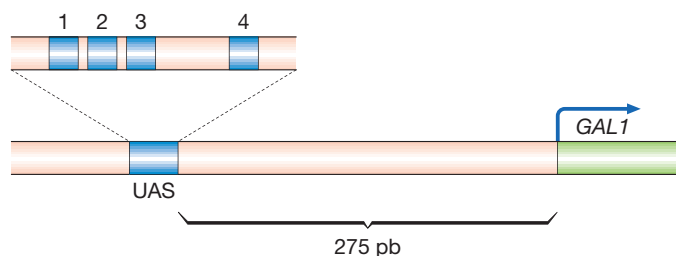


Figura 11.17 Struttura degli attivatori trascrizionali. Questa struttura schematica e semplificata dell'attivatore trascrizionale Gal4 mostra la presenza di due domini funzionali separati: uno che lega il DNA come un dimero, l'altro che attiva la trascrizione interagendo con il complesso di inizio e la polimerasi.

Figura 11.18 Sequenze regolatrici del gene di lievito *GAL1*. L'elemento UAS (*Upstream Activating Sequence*) contiene quattro siti di legame, lunghi ciascuno 17 pb, che legano specificamente il dominio di legame al DNA della proteina Gal4.



Un secondo esperimento ha permesso di dimostrare che il dominio di legame al DNA e il dominio di attivazione di un fattore trascrizionale sono distinti e, spesso, funzionalmente intercambiabili. Davanti al gene reporter *lacZ* è stata posta la sequenza di DNA riconosciuta dal repressore batterico LexA, che abbiamo già incontrato nel capitolo 7 (*vedi* finestra 7.1). Una proteina artificiale in cui il dominio di legame al DNA di Gal4 è stato sostituito dal dominio di legame al DNA di LexA può legarsi alla sequenza di DNA riconosciuta da LexA, ma non è in grado di attivare la trascrizione in quanto priva del dominio di attivazione trascrizionale. Se, però, si riesce a far produrre, tramite le tecnologie del DNA ricombinante (*vedi* cap. 21), una proteina ibrida contenente il dominio di legame al DNA di LexA fuso al dominio di attivazione trascrizionale di Gal4, la trascrizione viene attivata normalmente (**fig. 11.19B**). Questo esperimento in cui un repressore batterico è stato trasformato in un attivatore eucariotico ha dimostrato in maniera inequivocabile che gli attivatori della trascrizione sono composti da due domini separati. Il fatto che sia possibile separare il dominio di attivazione dal dominio di legame al DNA ha permesso di mettere a punto un metodo molto potente per lo studio d'interazioni proteina-proteina, conosciuto come **saggio del doppio ibrido** (*vedi* par. 20.12).

Come abbiamo detto, gli attivatori della trascrizione agiscono aumentando la velocità e il numero di trascritti prodotti da un determinato gene. Per ottenere questo, gli attivatori possono: (1) stimolare il reclutamento dei fattori basali della trascrizione di Pol II per formare il complesso di pre-inizio; (2) modificare i fattori di trascrizione sia a livello sterico che mediante modificazioni post-traduzionali (quale la fosforilazione, *vedi* par. 19.4); (3) interagire con il promotore del gene che deve essere trascritto e richiamare i complessi di rimodellamento e modificazione della cromatina, per liberare la regione del promotore dai nucleosomi e permettere la formazione del complesso d'inizio.

Tutte queste funzioni si realizzano, principalmente, tramite interazioni proteina-proteina tra gli attivatori e i loro specifici interattori, quali proteine che fanno parte del complesso del mediatore, la polimerasi stessa o altri fattori trascrizionali. Gli attivatori, inoltre, rispondono a stimoli esterni, che possono provocare modificazioni conformazionali nella proteina, rendendo così disponibile il dominio di attivazione. Molti fattori di trascrizione sono membri di famiglie proteiche, come per esempio i recettori degli ormoni steroidei, che hanno in comune le regioni di risposta all'ormone. In genere, queste proteine hanno strutture simili per quanto riguarda il dominio di legame al DNA, ma domini di attivazione diversi perché diverso è il gene che devono controllare e, quindi, il bersaglio specifico attraverso il quale questo controllo avviene. Gli attivatori trascrizionali sono, quindi, proteine modulari in cui domini con funzioni diverse vengono assemblati per formare molecole funzionali. Oltre ai domini di legame al DNA e al dominio di transattivazione, è presente in genere un dominio di dimerizzazione, che permette agli attivatori di legarsi al DNA come dimeri. Questa caratteristica amplifica la loro capacità di riconoscere sequenze specifiche sul DNA. Se l'attivatore funziona in risposta a degli stimoli, è presente anche un dominio che gli permette di legare il "ligando" specifico per l'attivazione del gene che è sotto il controllo dello stimolo.

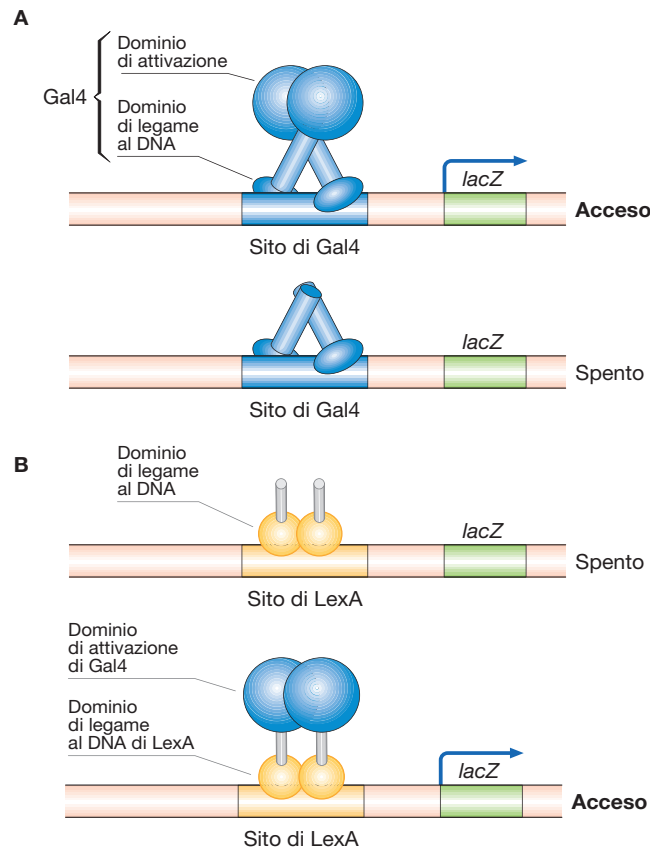


Figura 11.19 Esperimento dello scambio dei domini (*domain swapping*). Con questo esperimento è stata dimostrata la struttura degli attivatori degli eucarioti costituiti da due domini separati, quello di legame al DNA e quello di attivazione. In (A) la proteina Gal4 è in grado di legarsi al DNA e attivare la trascrizione; ma se viene privata del dominio di attivazione, la trascrizione non avviene. In (B) il dominio di legame al DNA di LexA, che è un repressore batterico, è stato fuso al dominio di attivazione di Gal4. Come si vede, la trascrizione è attivata solo se a monte del promotore è presente la sequenza di riconoscimento specifica per LexA. In questo esperimento è stato usato a valle del promotore il gene re-

porter *lacZ*, che permette la quantificazione del livello di trascrizione, misurando l'attività della β -galattosidasi. Infatti, questa è facilmente misurabile perché le colonie di lievito in cui sono presenti i vettori che contengono il gene reporter e le varianti di Gal4 da studiare, diventano di colore blu se il gene *lacZ* viene espresso. La colorazione blu è dovuta alla presenza della sostanza X-gal nelle piastre di agar dove vengono fatti crescere i lieviti trasformati (vedi cap. 21). X-gal è formato da un galattoside legato a un indolo e, se il gene *lacZ* è trascritto, la β -galattosidasi prodotta separa il galattoside dall'indolo, che venendo ossidato determina la caratteristica colorazione blu delle colonie.

Poiché gli attivatori agiscono, normalmente, sul DNA, contengono anche una sequenza amminoacidica che permette loro di localizzarsi nel nucleo (NLS, *Nuclear Localization Signal*). In alcuni casi, possono avere anche dei segnali di trasporto dal nucleo al citoplasma che regolano la loro funzione.

11.7 Domini funzionali dei transattivatori: domini di legame al DNA

Vediamo ora quali sono le strutture proteiche più frequentemente identificate nei domini di legame al DNA degli attivatori trascrizionali. Analogamente a quanto osservato nei repressori batterici, questi domini sono costituiti, nella maggior parte dei casi, da strutture ad α -elica che riconoscono la sequenza specifica di DNA interagendo con il solco maggiore della doppia elica. Tale solco, avendo più atomi in grado di ricevere o donare legami idrogeno, ha un "linguaggio" più ricco del solco minore per formare legami ad alta affinità con i residui amminoacidici che si affacciano sull' α -elica. Le interazioni specifiche tra DNA e proteina sono dovute principalmente a legami idrogeno e a interazioni

idrofobiche che la proteina forma con la sequenza di DNA. È stato osservato che il riconoscimento coinvolge in genere sequenze di 3-7 pb sul DNA, mentre, per poter essere unica nella vasta complessità di sequenze di un genoma eucariotico, una sequenza nucleotidica di riconoscimento dovrebbe essere lunga almeno 16 pb. In effetti la specificità di legame è aumentata dal fatto che per lo più gli attivatori dimerizzano e, quindi, le sequenze di riconoscimento sono costituite da elementi doppi sul DNA. Come vedremo, tale proprietà aumenta enormemente la possibilità di controllo combinatoriale della trascrizione.

La cristallografia a raggi X ha risolto negli ultimi vent'anni un numero molto grande di strutture di proteine che legano il DNA e confrontando questi domini, gli attivatori sono stati raggruppati in diverse categorie proprio in base ai "motivi" con i quali si legano al DNA.

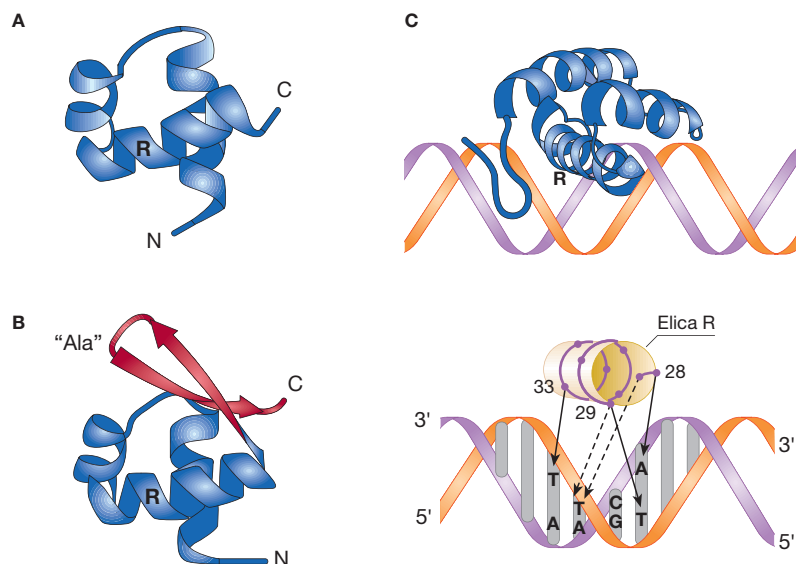
Elica-giro-elica (*helix-turn-helix*) e omeodominio

Il **motivo elica-giro-elica** (*helix-turn-helix*, **HTH**) è stato uno dei primi a essere individuato negli studi sui repressori batterici. Abbiamo già incontrato questa struttura nel repressore dell'operone *lac* e nel repressore Cro del batteriofago lambda. La **figura 11.20A** mostra la struttura generica del motivo HTH, che è costituita da tre α -eliche ripiegate in cui l'elica 3, chiamata anche **elica R (elica di Riconoscimento)**, si posiziona nel solco maggiore del DNA. I residui amminioacidici dell'elica R che si affacciano sul DNA interagiscono specificamente con le basi della sequenza di riconoscimento e, come atteso, la sostituzione di una di queste basi abbassa notevolmente l'affinità della proteina per quella sequenza.

Esistono numerose varianti della struttura HTH. Per esempio, in alcuni casi le α -eliche possono essere due o anche quattro, sempre con un'elica R che interagisce nel solco maggiore della doppia elica. Un'altra variante, che si trova in vari attivatori ed è mostrata in **figura 11.20B**, è rappresentata dalla struttura **HTH winged** che presenta un'"ala" (*wing*) formata da due foglietti β .

Di questa famiglia di attivatori fa parte anche l'**omeodominio (HD)**. Questo motivo, costituito da circa 60 amminoacidi, è stato individuato per la prima volta in *Drosophila* ed è caratteristico dei regolatori dei geni omeotici che controllano lo sviluppo dell'intero organismo. È presente anche in regolatori trascrizionali di mammiferi, incluso l'uomo, che controllano più di un centinaio di geni. L'omeodominio (**fig. 11.20C**) è costituito da tre α -eliche strutturate come nel dominio

Figura 11.20 Dominio "elica-giro-elica" (HTH) e omeodominio. (A) Il dominio HTH è costituito in genere da tre (alle volte due o anche quattro) α -eliche una delle quali, chiamata elica R (riconoscimento), si posiziona all'interno del solco maggiore della doppia elica di DNA. (B) Una variante del dominio HTH, chiamata HTH winged, è caratterizzata dalla presenza di un'"ala" (in rosso) formata da due foglietti β . (C) L'omeodominio è un caso particolare di motivo HTH che contiene un'estensione flessibile non strutturata che va a interagire con il solco minore del DNA.



HTH, in cui l'elica R contatta il solco maggiore riconoscendo una sequenza di 6 pb. L'omeodominio contiene, in aggiunta alle tre eliche, anche un braccio flessibile, non strutturato, che interagisce specificamente con il solco minore del DNA, come mostrato in **figura 11.20C**. L'affinità di legame è data dalla combinazione dei due tipi di riconoscimento; comunque le interazioni con il solco minore sono essenziali per la specificità e la stabilità del legame.

Dominio a dita di zinco (*zinc finger*)

Il **motivo a dito di zinco** è stato individuato per la prima volta nel fattore TFIIIA di *Xenopus laevis*, necessario per la trascrizione dell'rRNA 5S. L'espressione "dito di zinco" deriva dallo schema della sua struttura, dove un atomo di zinco coordina quattro residui di cisteine e/o istidine, formando un'ansa che ricorda la forma di un dito. Nell'esempio mostrato in **figura 11.21A** la sequenza amminoacidica presenta, a livello di struttura primaria, due cisteine e due istidine distanziate tra loro di 3, 13 e 3 residui. Nella struttura secondaria i quattro amminoacidi (due cisteine e due istidine) vanno a legare un atomo di zinco, formando quindi un'ansa di circa 12-13 amminoacidi. La risoluzione della struttura tridimensionale di questo motivo ha rivelato che l'ansa è divisa in due domini diversi formando, a sinistra, una struttura a foglietto β e, a destra, un' α -elica. La struttura secondaria e quella tridimensionale di un dominio a dito di zinco sono mostrate in **figura 11.21B e C**, dove l'atomo di zinco è legato covalentemente alle due cisteine e alle due istidine, secondo lo schema $\text{Cys}_2/\text{His}_2$. Un intero dito di zinco è generalmente formato da 23 residui con le cisteine e le istidine molto conservate alla sua base, anche se, va detto, esistono varianti con schemi diversi, per esempio quattro Cys. La punta dell'ansa è formata in genere da amminoacidi idrofobici che stabilizzano la struttura dell'ansa, mentre l' α -elica va a interagire con il solco maggiore della doppia elica del DNA; linker di 7-8 amminoacidi separano più dita adiacenti. Questa struttura del dominio di legame al DNA è diffusa in molte famiglie di attivatori trascrizionali e, in generale, si presenta con diverse ripetizioni della struttura a dita (**fig. 11.21D**): ogni dito aggiuntivo aumenta la specificità e l'affinità di legame al DNA. Alcuni attivatori di *Drosophila* contengono 2 dita di zinco, il fattore Sp1, che agisce su numerosi geni, ne contiene 3, mentre il fattore TFIIIA ne contiene fino a 9.

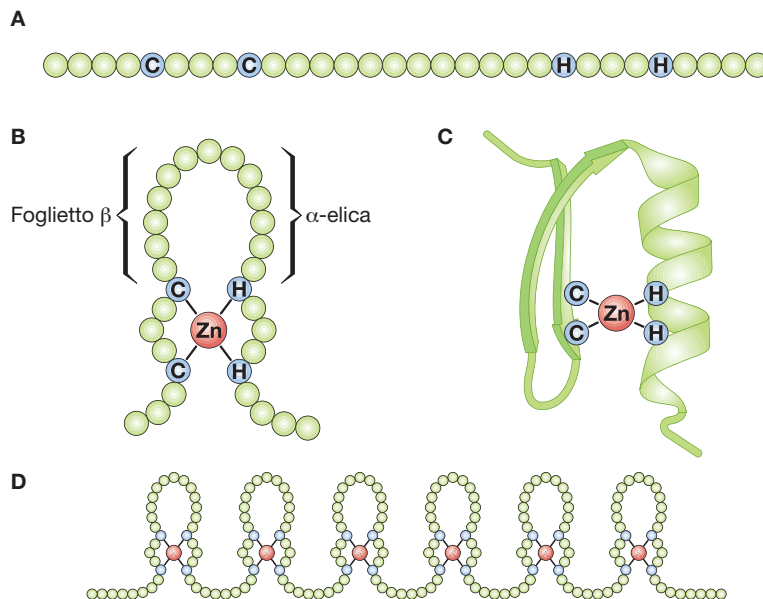
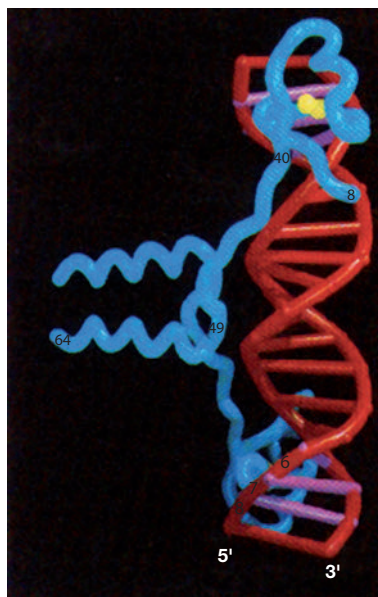


Figura 11.21 Dominio a dita di zinco. (A) A livello di struttura primaria si osserva la presenza di 2 cisteine (C) e 2 istidine (H), o in alcuni casi di 4 cisteine, spaziate rispettivamente di circa 3, 12 e 3 residui amminoacidici. (B) Questi quattro amminoacidi legano un atomo di zinco, facendo assumere alla regione la struttura a "dito". (C) La regione a dito presenta due metà distinte: un foglietto β e una regione ad α -elica che lega il DNA. (D) In genere più dita di zinco sono presenti in serie nelle proteine che legano il DNA.

Figura 11.22 L'attivatore del lievito Gal4 si lega al DNA come un dimero attraverso due domini a dita di zinco che riconoscono la sequenza specifica di 17 pb. Come mostrato in figura, gli ultimi 164 amminoacidi costituiscono il sito di legame al DNA e il sito di dimerizzazione, attraverso la formazione di due strutture coiled-coil, perpendicolari al dominio di legame al DNA. [Fonte: Marmorstein, R., Carey, M., Ptashne, M., Harrison, S.C. (1992), *Nature*, 356(6368), pp. 408-414.]



La **figura 11.22** mostra l'attivatore di lievito Gal4, che, come abbiamo visto precedentemente, attiva il gene *GALI*; Gal4 si lega al DNA come un dimero tramite due dita di zinco che riconoscono elementi di 17 pb. Dimeri di dita di zinco caratterizzano anche la grande famiglia di attivatori rappresentati dai recettori degli ormoni steroidei. Questi dimerizzano e riconoscono il loro sito di legame sul DNA costituito da due sequenze di 6 nt orientate in direzione opposta, con una spaziatura di 3 nt tra loro. In queste proteine, che hanno dita di zinco del tipo Cys₂/Cys₂, non si riscontrano ripetizioni in serie della struttura a dito. Per esempio, i recettori del glucocorticoide e dell'estrogeno hanno ciascuno due dita, in arrangiamento Cys₂/Cys₂, le cui α -eliche legano il DNA nel solco maggiore. Nel dimero, le due α -eliche si trovano sulla stessa faccia del DNA e si legano a due giri di doppia elica adiacenti. Le sequenze dei recettori degli estrogeni e del glucocorticoide sono molto conservate e le differenze si trovano soprattutto alla base delle dita. Come vedremo in seguito, il legame al DNA e la successiva attivazione della trascrizione dipendono dal legame con l'ormone, che induce dei cambiamenti conformazionali nel recettore, attivandolo.

Domini a cerniere di leucina (*leucine zipper*)

Numerosi attivatori trascrizionali legano il DNA grazie al motivo a **cerniera di leucine** (*leucine zipper*). Questi attivatori sono dimeri formati da due lunghe strutture ad α -elica, ciascuna delle quali contiene, contemporaneamente, un dominio di dimerizzazione e un dominio basico di legame al DNA. Quest'ultimo è formato dalla parte terminale delle due α -eliche che si inseriscono nel solco maggiore della doppia elica ma sulle facce opposte del DNA, dove gli elementi di risposta sono costituiti da una coppia di sequenze di 4 pb con orientamento opposto, distanziati da un nucleotide.

Il dominio di dimerizzazione presenta, a livello di struttura primaria, un residuo di leucina ogni 7 amminoacidi (**fig. 11.23A**). In questo modo, una volta acquisita la struttura ad α -elica, e poiché un giro di elica corrisponde a circa 3,5 residui amminoacidici, le leucine vengono a trovarsi, ogni due giri, sullo stesso lato dell'elica (**figura 11.23B**). L'assemblaggio della struttura dimerica è dovuto ai legami idrofobici tra le leucine che protrudono dalle due α -eliche proprio come i denti di una cerniera lampo. Questa struttura dimerica, mostrata in

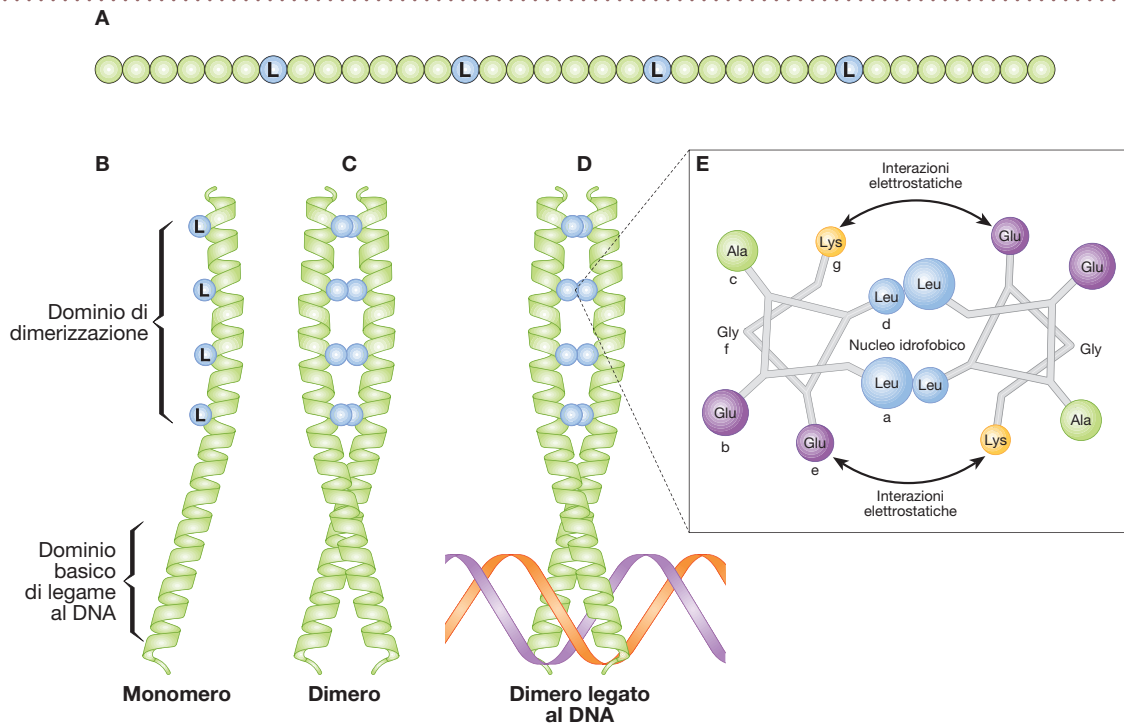


Figura 11.23 Dominio a cerniera di leucine. (A) A livello di struttura primaria si osserva la presenza di più leucine (L) spaziate a 7 amminoacidi di distanza l'una dall'altra. (B) La struttura ad α -elica della regione fa sì che le leucine vengano a trovarsi, a giri alterni, su una faccia dell' α -elica. (C) L'interazione idrofobica delle leucine di

due monomeri risulta nella formazione di un dimer. (D) Tramite il dominio basico il dimer si lega al DNA, nel solco maggiore. (E) Interazione idrofobica tra le leucine esposte su una faccia di ciascuna delle due α -eliche.

figura 11.23C, è in grado di formare un'interazione stabile con la doppia elica del DNA (**fig. 11.23D**). In **figura 11.23E** è infine evidenziato il nucleo idrofobico responsabile della dimerizzazione.

È importante notare che, poiché la dimerizzazione mediata dalle leucine non è specifica, si possono ottenere sia omodimeri che eterodimeri tra diversi tipi di α -eliche contenenti leucine, ampliando così il numero di possibili combinazioni tra domini di attivazione e domini di legame al DNA. Le **figure 11.24A** e **11.24B** mostrano come i fattori omodimerici 1/1 e 2/2 riconoscono sequenze di DNA diverse che contengono gli elementi invertiti di 4 pb, mentre l'attivatore eterodimerico 1/2 (**fig. 11.24C**) riconosce una sequenza formata dalla combinazione dei due mezzi siti riconosciuti dai monomeri 1 e 2.

Il motivo a cerniera di leucine è stato descritto per la prima volta nella proteina C/EBP che lega la CAAT box presente in molti enhancer; appartiene a questo gruppo di attivatori la proteina AP-1, che è un eterodimero formato dall'unione dei proto-oncogeni c-Jun e c-Fos e regola l'espressione genica in risposta a vari stimoli, quali stress, infezioni virali e batteriche, e l'attivazione da parte di citochine. Mentre c-Jun può formare omodimeri ed eterodimeri, c-Fos può formare solo eterodimeri. In lievito, il fattore di trascrizione Gcn4, che controlla la trascrizione di molti geni per la biosintesi degli amminoacidi, appartiene anch'esso alla famiglia dei leucine zipper ed è formato da due α -eliche che hanno il dominio di legame al DNA ricco di residui basici, come arginine e lisine.

Dominio elica-ansa-elica (*helix-loop-helix*)

Il **dominio elica-ansa-elica** (*helix-loop-helix*, HLH), la cui struttura monomerica è mostrata in **figura 11.25A**, è caratterizzato da due α -eliche anfipatiche (cioè

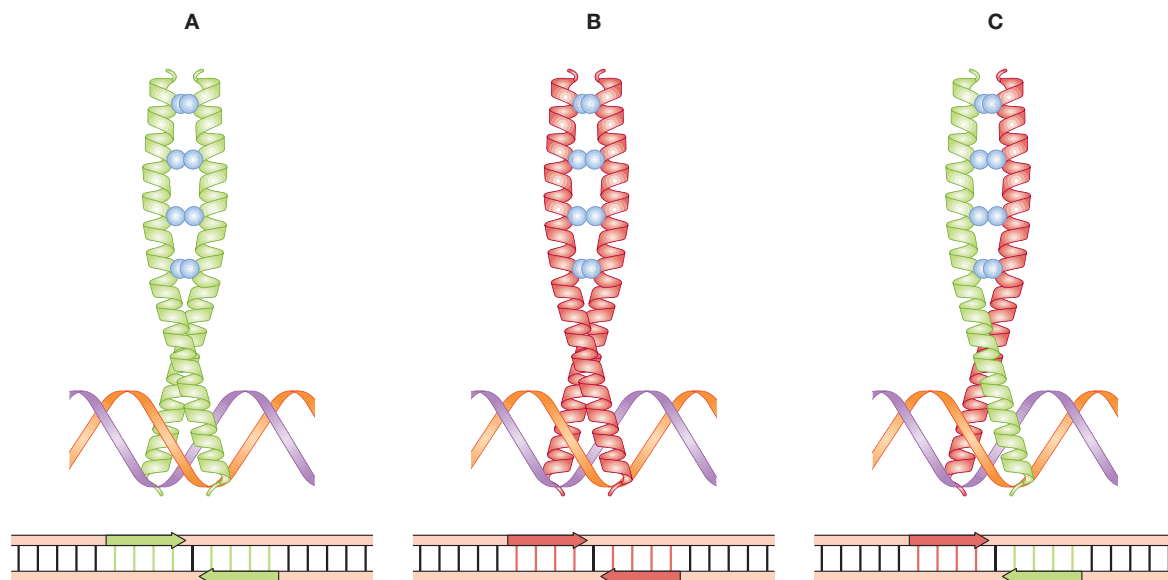


Figura 11.24 Dimerizzazione e controllo combinatoriale. La dimerizzazione mediata da cerniere di leucina non è specifica, per cui molti di questi attivatori possono formare eterodimeri. (A) e (B) Due omodimeri, rispettivamente in verde e in rosso, si legano

alla propria sequenza di riconoscimento sul DNA. (C) Un eterodimero formato da due diversi monomeri (uno in verde, l'altro in rosso) si lega a un sito di riconoscimento sul DNA formato da due mezzi siti, specifici ciascuno per un monomero.

presentano residui idrofobici su una faccia e residui carichi sull'altra) connesse da un'ansa. In generale un monomero è costituito da circa 40 amminoacidi e l'elica più lunga contiene, nella regione N-terminale, un dominio basico che le permette di interagire con il DNA; infatti, questi attivatori sono anche chiamati **BHLH** (*basic helix-loop-helix*). Come mostrato nella **figura 11.25B e C**, la formazione di un dimero, che rappresenta la struttura in grado di legare il DNA, avviene mediante interazioni tra le facce idrofobiche dei due monomeri.

In realtà, i fattori che hanno questo motivo, sono in genere eterodimerici. Anche in questo caso, come nei motivi a cerniera di leucine, la formazione di eterodimeri aumenta la complessità combinatoriale di questi attivatori. In genere, riconoscono una sequenza sul promotore, chiamata **E-box**, che spesso contiene una sequenza palindromica canonica, tipo CANNTG. A questa famiglia di proteine appartengono molti attivatori che hanno un ruolo nello sviluppo, come, per esempio, Myo-D, che svolge una funzione fondamentale nel differenziamento delle cellule muscolari. Un altro esempio è la proteina c-Myc, che è coinvolta nella trasformazione tumorale, per esempio nel linfoma di Burkitt. Come abbiamo detto, spesso questi fattori formano degli eterodimeri e la loro attività è regolata dalla dimerizzazione: una subunità può essere controllata dalla richiesta fisiologica, mentre l'altra è espressa costitutivamente. Per esempio, un fattore importante per la regolazione del ritmo circadiano, cBMAL1-CLOCK, è un eterodimero BHLH.

Controllo combinatoriale della trascrizione

Come abbiamo visto, gli elementi enhancer e quelli contenuti nella regione prossimale dei promotori sono implicati nel legame degli attivatori e la loro disposizione è alla base del controllo combinatoriale della trascrizione. Questa varietà di sequenze corrisponde, infatti, alla varietà sia numerica che strutturale degli attivatori. Le combinazioni che possono insorgere con diverse sequenze di riconoscimento, e con la dimerizzazione di diversi tipi di monomeri, sono molto numerose. Per esempio, tre fattori e tre sequenze di riconoscimento possono dar vita a sei possibili complessi di attivazione; se uno dei tre fattori considerati può

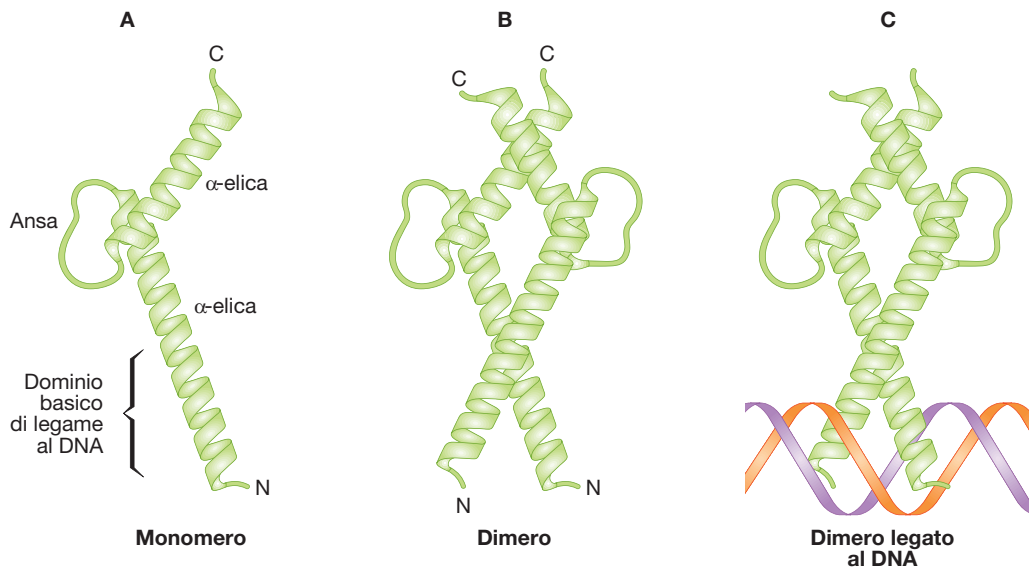


Figura 11.25 Dominio “elica-ansa-elica basico” (BHLH). (A) Il monomero di un fattore del tipo HLH basico è costituito da due α -eliche connesse da un’ansa; l’elica più lunga presenta una porzione basica che rappresenta il dominio di legame al DNA. (B) Due

monomeri interagiscono, tramite la prima α -elica, a formare un dimer. (C) Il dimer interagisce, tramite il dominio basico, con il solco maggiore della doppia elica del DNA.

dimerizzare anche con un fattore di inibizione, verrà persa la capacità di attivare la trascrizione su un certo numero di siti. Se si immagina uno schema simile su una sequenza enhancer che contiene anche sei o sette siti di legame, si capisce come un controllo combinatoriale amplifichi il numero di possibilità, utilizzando un numero limitato di fattori.

Per fare un esempio riportiamo in **figura 11.26A** la struttura tridimensionale dell’enhanceosoma dell’interferone- β , con i fattori di trascrizione legati ai diversi siti di legame sul DNA. Come si può osservare, l’intera sequenza di DNA dell’enhancer contenente molti siti contigui è completamente ricoperta dagli attivatori. L’associazione di 8 proteine con l’enhancer crea una superficie continua che riconosce circa 50 pb sul DNA. I vari siti di legame per gli attivatori si sovrappongono e le variazioni conformazionali in alcuni di essi contribuiscono ad aumentare la specificità di legame e il posizionamento dei fattori. È l’intera sequenza nucleotidica a formare l’elemento di regolazione e questo spiega anche perché tale sequenza sia estremamente conservata in tutti i mammiferi. Possiamo riconoscere, in questa struttura, alcuni dei motivi d’interazione con il DNA che abbiamo sopra descritto, come la cerniera di leucine nel caso di Jun e Atf e il motivo elica-giro-elica per IRF-3.

La superficie continua dei vari attivatori, mostrata in **figura 11.26B**, fornisce molti elementi di riconoscimento ed è la combinazione di questi segnali che rende specifico il legame con il mediatore, i TAF e la polimerasi. Infatti, alcuni attivatori come Jun-Atf, Nfk-B e molti altri sono coinvolti nella regolazione di numerosi geni. È quindi la combinazione tra diversi attivatori che rende molto specifico il meccanismo di regolazione.

Come abbiamo visto, sono state risolte le strutture tridimensionali dei domini di legame al DNA di molti fattori trascrizionali; ciò nonostante, i meccanismi molecolari precisi attraverso i quali si realizza l’attivazione trascrizionale sono ancora in gran parte sconosciuti. Sappiamo che molti domini di attivazione trascrizionale hanno regioni ricche di amminoacidi acidi e che mutazioni che sostituiscono uno di questi amminoacidi con altri portano, in molti casi, all’abolizione dell’attivazione, mentre mutazioni che aumentano generalmente la carica acida del dominio aumentano l’intensità di attivazione. Ancora una volta la proteina Gal4 è stata tra

le prime a suggerire che la carica acida dei domini d'attivazione fosse importante per la loro funzione. Quindi, ritornando alla **figura 11.26B**, possiamo concludere che è molto probabile che l'intera "superficie" che tutti gli attivatori contribuiscono a formare, definisca una regione con la capacità di formare specifiche interazioni proteina-proteina con gli altri fattori che formeranno il complesso di pre-inizio. Per attivare un gene, quindi, la specificità è data dalla regione di legame al DNA dei fattori trascrizionali che si legano sull'enhancer solo quando le condizioni fisiologiche della cellula lo richiedono e la conformazione della cromatina lo permette. Il domino di regolazione globale dell'enhanceosoma è rappresentato dalla somma dei segnali che i singoli attivatori contribuiscono a formare, per promuovere interazioni forti con il mediatore e i fattori basali di trascrizione.

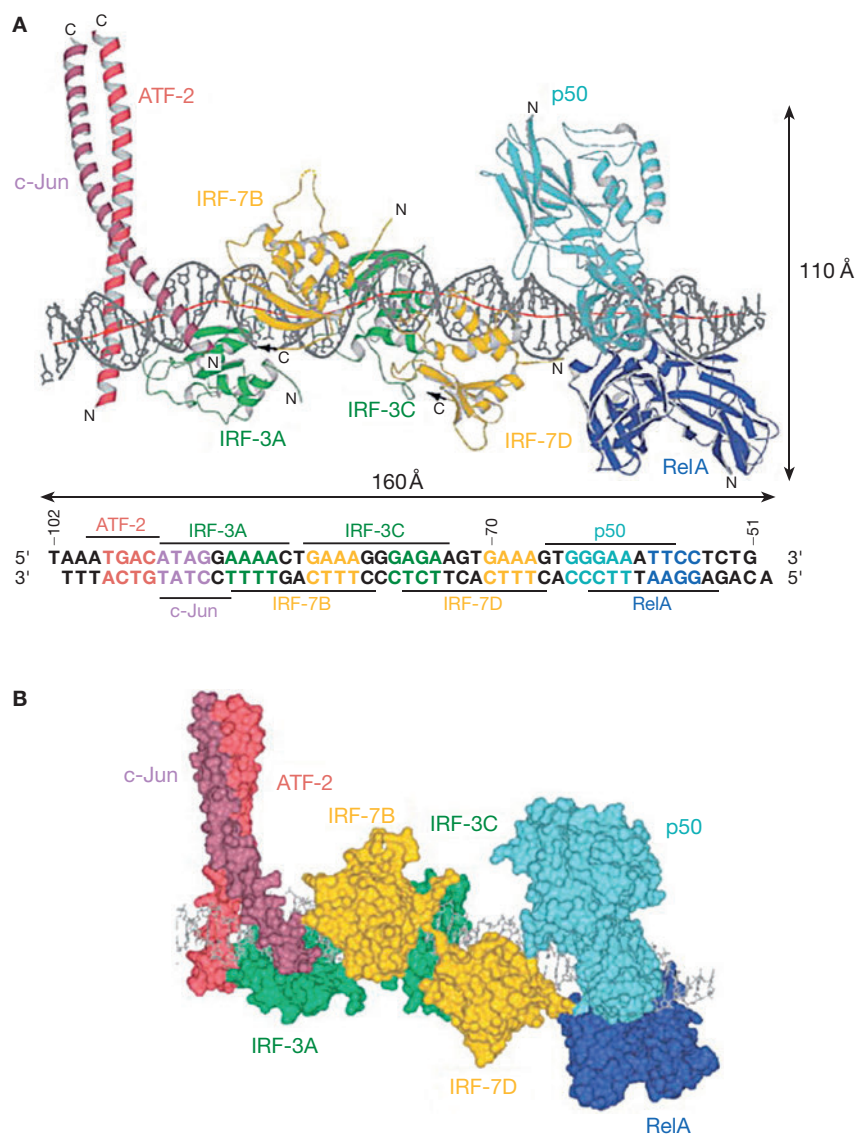


Figura 11.26 Struttura tridimensionale dell'enhaneosoma dell'interferone- β , con diversi fattori di trascrizione legati a loro siti di legame sul DNA. (A) L'intera sequenza di DNA dell'enhancer, contenente molti siti contigui, è completamente ricoperta da attivatori. Si possono riconoscere vari motivi d'interazione con il DNA, come la cerniera di leucine nel caso di Jun e ATF e il motivo elica-giro-elica per IRF-3. (B) La rappresentazione ad atomi pieni

(superficie delle proteine) ci fa vedere che l'insieme degli attivatori forma un complesso continuo che può essere riconosciuto dai fattori che fanno parte del complesso d'inizio. Infatti, alcuni attivatori come Jun-Atf, Nfk-B e molti altri sono coinvolti nella regolazione di numerosi geni. È, quindi, la combinazione tra diversi attivatori che rende specifico il meccanismo di regolazione. [Fonte: Panne, D., Maniatis, T. e Harrison, S.C., (2007), *Cell*, 129(6), pp. 1111-1123.]

L'aspetto combinatorio della formazione di un complesso di fattori che attivano la trascrizione, suggerisce che gli attivatori agiscano attraverso reti di interazioni complesse, come descritto nella **finestra 11.2**.

Reclutamento del complesso di trascrizione

Nei batteri, il ruolo degli attivatori implicati nel controllo positivo della trascrizione è fondamentale per reclutare la RNA polimerasi sul promotore. Questi attivatori, come il repressore CI e CAP, interagiscono direttamente con la polimerasi e sono soprattutto necessari per quei promotori in cui la subunità σ da sola non è sufficiente a formare il complesso d'inizio. Anche negli eucarioti gli attivatori "reclutano" la polimerasi sul promotore, ma non lo fanno in modo diretto; essi agiscono indirettamente, tramite il mediatore e i TAF. La **figura 11.27A** mostra come un attivatore legato a un enhancer possa contattare il mediatore che è già legato alla polimerasi e il fattore basale TFIID, facilitando il loro legame al promotore. Un esperimento di bypass dell'attivatore, analogo a quello descritto nel capitolo 10 per i batteri (*vedi* par. 10.2 e fig. 10.14), ha dimostrato che in lievito, se si forza il mediatore a legarsi direttamente al DNA, può essere aggirato il ruolo di reclutamento di un attivatore. Utilizzando come sistema modello il promotore del gene *GAL1* di *Saccharomyces cerevisiae*, il sito di legame sul DNA riconosciuto da Gal4 è stato sostituito con la sequenza di DNA riconosciuta da LexA. La creazione di una proteina di fusione tra Gal11 (che fa parte del mediatore) e il dominio di legame al DNA di LexA, permette il reclutamento sul promotore della polimerasi, del mediatore, di TFIID e di vari TAF in assenza dell'attivatore Gal4 (**fig. 11.27B**). Questo esperimento dimostra che il ruolo degli attivatori consiste nel favorire il reclutamento del complesso di pre-inizio sul promotore. La specificità del sistema, come indicato precedentemente, è delegata alla specificità con la quale gli attivatori si legano al DNA.

Attivatori e cromatina

Probabilmente la maggiore differenza tra trascrizione negli eucarioti e nei batteri è legata al fatto che il DNA eucariotico è organizzato in una struttura cromatinica, mentre nei batteri è sempre sostanzialmente disponibile al legame dell'RNA

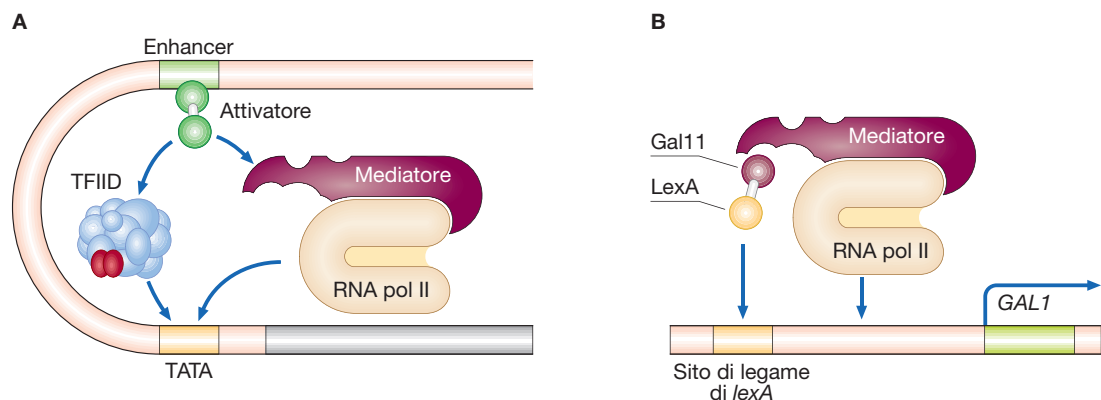


Figura 11.27 Attivazione della trascrizione mediante il reclutamento del complesso d'inizio e ruolo del mediatore. (A) Attivazione della trascrizione mediante il reclutamento del complesso di inizio; il mediatore legato alla polimerasi è in grado di riconoscere i fattori basali della trascrizione come il complesso di TFIID e legarsi al promotore. Gli attivatori, legati agli enhancer, riconoscono e legano in modo specifico alcuni componenti del mediatore. Quando queste interazioni sono stabili, il complesso è in grado di

trascrivere. (B) Esperimento di by-pass dell'attivatore. In questo esperimento si chiarisce che il ruolo degli attivatori legati all'enhancer è quello di reclutare l'oloenzima, formato dal mediatore e dalla polimerasi, sul promotore. Il dominio di legame al DNA di LexA è fuso alla proteina Gal11 che fa parte del complesso del mediatore. Se a monte del gene *GAL1* viene messa la sequenza di riconoscimento di LexA, la trascrizione viene attivata anche in assenza dell'attivatore specifico Gal4.

FINESTRA 11.2 I FATTORI DI TRASCRIZIONE AGISCONO ATTRAVERSO UNA VASTA RETE DI INTERAZIONI

Come abbiamo visto nel testo, i fattori trascrizionali degli eucarioti compiono la loro funzione di attivazione o repressione della trascrizione legandosi a brevi e talora estremamente degenerate sequenze sul DNA. Molte di queste sequenze possono essere riconosciute da più fattori ed essere localizzate in corrispondenza di geni diversi. Molti studi sulla specificità del legame dei fattori sul DNA sono stati fatti *in vitro*, utilizzando DNA e fattori purificati. Tuttavia *in vivo* il problema è più complesso, data la grande diversità tra cellule e tessuti di un organismo, in cui anche la sola posizione che una cellula assume all'interno di un tessuto la può rendere diversa dalla cellula adiacente. I fattori di trascrizione inoltre agiscono spesso in una maniera combinatoria, legando il DNA solo quando co-adiuvati da altri fattori.

Se si osserva la figura 11.26, che descrive gli attivatori legati all'enhancer del gene dell'interferone- β , il fattore cJun-ATF2 ha come bersaglio molti geni diversi. Il ruolo quindi dei fattori di trascrizione *in vivo* può essere potenzialmente diverso in ogni cellula. Se si considera che i fattori interagiscono tra loro e interagiscono con diverse sequenze di DNA, si può descrivere la loro azione come un network che, mettendo in relazione le varie interazioni, possa dare informazioni più precise sui possibili modi di agire e ai principi che sottendono alla regolazione dei geni negli eucarioti.

Risultati recenti suggeriscono che i network della trascrizione negli eucarioti si possono descrivere meglio come "continui" in cui ciascun fattore modula l'espressione di tutti i suoi geni target, anche migliaia, legandosi alle corrispondenti regioni regolatorie. I geni i cui elementi di controllo (ad es. promotore) sono occupati da un fattore di trascrizione in maniera elevata sono probabilmente i suoi target più specifici. D'altra parte, i fattori di trascrizione sono espressi nell'ordine delle decine di migliaia in una cellula e, come conseguenza, gli elementi in *cis* sul DNA a cui i fattori si legano sono occupati ad alto livello da fattori che hanno una funzione biologica di

attivazione o repressione dei geni corrispondenti, e possono comunque essere occupati a livello basso da quei fattori che sugli stessi geni non hanno una funzione regolatoria diretta. Per studiare *in vivo*, negli organismi, il ruolo dei fattori di trascrizione, identificare le regioni genomiche a cui si legano e il livello quantitativo di legame a queste regioni, recentemente è stata introdotta la tecnica di "immunoprecipitazione della cromatina associata al sequenziamento del DNA di nuova generazione" detta anche **ChIP-Seq**, descritta nel paragrafo 20.8. Questa nuova metodologia associa l'immunoprecipitazione con anticorpi specifici per il fattore di trascrizione di interesse con il sequenziamento di tutti i tratti di DNA legati *in vivo* sul genoma dal fattore di trascrizione stesso.

Il DNA immunoprecipitato viene purificato e successivamente sequenziato su larga scala mediante le piattaforme NGS (vedi par. 20.8). L'analisi dei dati ottenuti dal sequenziamento, che consiste solitamente nella mappatura o allineamento delle sequenze ottenute sul genoma corrispondente, mostrerà un arricchimento di sequenze provenienti dalle regioni immunoprecipitate, graficamente visualizzabili come picchi (fig. 11F2.1), corrispondenti alle regioni di DNA dove quel fattore era legato nel sistema cellulare utilizzato. Questo processo sperimentale rende possibile non solo l'individuazione dei siti di legame nel genoma intero, ma anche la stima dei livelli quantitativi del legame. Usando questa tecnica sono state generate migliaia di mappe di questo tipo per diversi fattori di trascrizione, in diverse specie e tipi cellulari. Per esempio, il gruppo di Mark Biggin del Lawrence Berkeley Laboratory, analizzando alcuni fattori di trascrizione di *Drosophila melanogaster*, ha trovato che questi si legavano in migliaia di posizioni differenti sul genoma, regolando potenzialmente quasi la metà dei geni che codificano per proteine. L'occupazione delle regioni era però quantitativamente differente e solo i legami con livelli più alti erano biologicamente significativi.

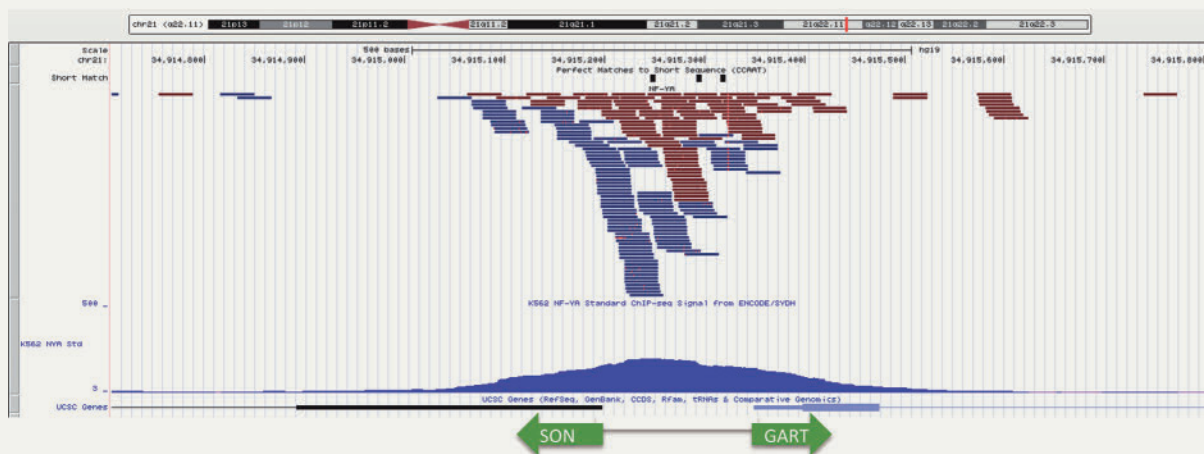


Figura 11F2.1 Il legame della subunità A della proteina NFY al promotore bidirezionale dei geni SON/GART sul chr21 umano risultante da un esperimento di ChIP-Seq. Si vedono in alto i read NGS mappati sul genoma (il colore blu e rosso si riferisce al loro

orientamento forward e reverse, rispettivamente) e il plot dell'arricchimento in basso. Poiché NFY lega la CCAAT box, si noti come in corrispondenza del picco si trovano proprio gli oligo CCAAT sul DNA. (Per gentile concessione di Giulio Pavesi, Università di Milano.)

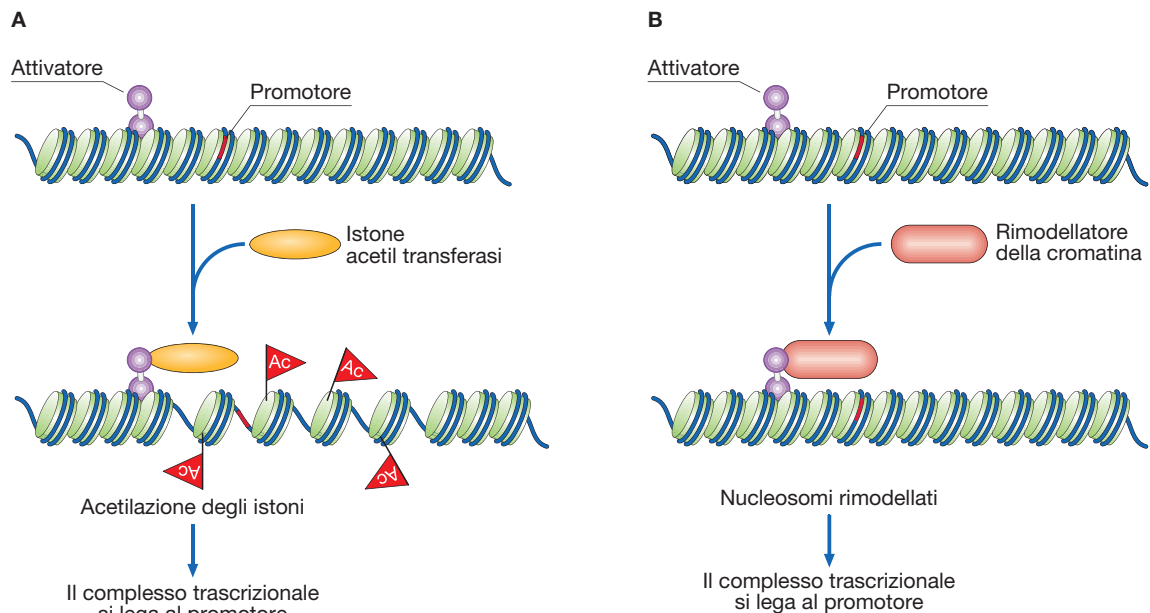


Figura 11.28 Ruolo degli attivatori nella regolazione della struttura della cromatina. Oltre a interagire con il mediatore e reclutare il complesso di inizio, alcuni attivatori hanno un ruolo attivo nella regolazione della cromatina. (A) Un promotore è inaccessibile a causa della struttura compatta della cromatina. Un attivatore specifico recluta sulla regione una istone acetil transferasi,

che rende meno compatta la cromatina modificando le code N-terminali degli istoni. (B) Il promotore diventa accessibile quando un attivatore specifico recluta nella regione un rimodellatore della cromatina che, modificando la struttura della cromatina (nucleosomi gialli), rende disponibile il promotore per la formazione del complesso di inizio.

polimerasi. Il ruolo di molti attivatori è di attivare la trascrizione di geni con una struttura della cromatina compatta, reclutando complessi di modificazione o rimodellamento della cromatina stessa. Poiché l'organizzazione cromatinica locale di una regione genomica è spesso ereditabile, cioè trasmessa dalla cellula madre alle cellule figlie, si fa rientrare questi meccanismi regolativi nell'ambito dei processi epigenetici di cui parleremo più in dettaglio nel capitolo 17 sulle regolazioni epigenetiche, a cui rimandiamo (*vedi* par. 17.2 e figg. 17.5 e 17.6). Per ora ricordiamo solo, brevemente, che un promotore completamente inaccessibile alla RNA polimerasi e al complesso di pre-inizio può essere attivato dal legame di un fattore specifico a un sito che si trovi a monte. Questo sito deve trovarsi sul lato esterno del nucleosoma e in grado quindi di legare proteine. Come mostrato in **figura 11.28A**, questo fattore specifico recluta sulla regione una istone acetiltransferasi (HAT) che, acetilando le code degli istoni contigui, promuove l'apertura di tutta la regione del promotore, dove si può così formare il complesso di pre-inizio. Alternativamente (**fig. 11.28B**) il fattore specifico recluta sulla regione un complesso di rimodellamento della cromatina, o "rimodellatore", che rende la zona del promotore accessibile al complesso di trascrizione.

Alcuni attivatori della trascrizione che normalmente reclutano il complesso di pre-inizio sono anche in grado di reclutare complessi di modificazione della cromatina, come la proteina Gal4 di lievito, che in alcuni casi, recluta sul promotore il complesso SAGA (Spt-Ada-Gcn5-acetil transferasi) rendendo più accessibile la regione da trascrivere.

La **figura 11.29** riassume quanto abbiamo descritto sull'attivazione e la formazione del complesso di trascrizione. Nel pannello A è schematizzato come i vari attivatori, legati agli elementi prossimali e distali di un promotore, interagendo con co-attivatori, il mediatore e i vari TAF promuovono la formazione

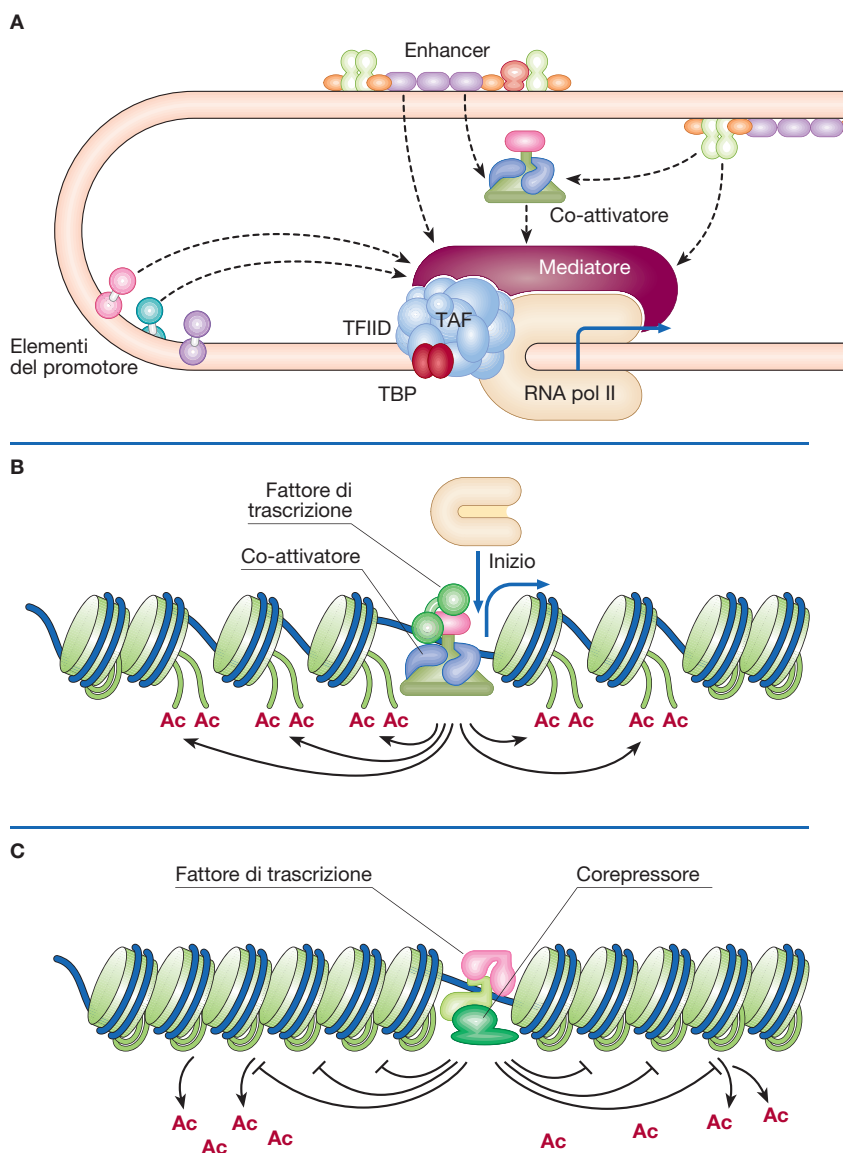


Figura 11.29 Visione globale dei meccanismi di attivazione della trascrizione. (A) I contatti tra gli attivatori, il mediatore e i TAF portano alla formazione di un complesso stabile che permette alti livelli di trascrizione. (B) Il reclutamento delle istone acetil transferasi

apre la regione della cromatina e permette la formazione del complesso di inizio. (C) Il reclutamento di un'attività di istone deacetilasi rende più compatta la cromatina e reprime la trascrizione rendendo inaccessibile il promotore.

del complesso di pre-inizio. Nel pannello B è mostrato come un attivatore si lega al DNA e recluta un co-attivatore che, modificando le code degli istoni, apre la cromatina; nel pannello C è mostrato come un fattore che agisce da repressore, reclutando sul DNA un fattore con attività di deacetilasi, rende più compatta e inaccessibile la cromatina. Infatti, analogamente a quanto accade per gli attivatori, esistono una serie di fattori che agiscono da repressori. Questi possono funzionare in vari modi:

- competendo per i siti di legame al DNA di un attivatore;
- mascherando, attraverso interazioni proteina-proteina, il dominio di attivazione;
- interagendo con il mediatore e bloccare il complesso sul promotore;
- reclutando delle deacetilasi, come mostrato in figura.

I repressori rappresentano un elemento molto importante della regolazione ge-

nica negli eucarioti; infatti, possono reprimere la trascrizione di un gene fino a quando uno stimolo esterno, per esempio inattivando il repressore, non dia via libera ai vari attivatori specifici per quel gene.

Attivazione e segnali

La regolazione genica negli eucarioti è legata alla capacità delle cellule di reagire a una grande varietà di stimoli e segnali esterni. Le cellule rispondono alla presenza di ormoni, citochine, fattori di crescita, gradienti di concentrazione di ioni, luce, contatti con altri tipi cellulari e così via, fino alla risposta ad agenti patogeni.

Queste sollecitazioni “dall'esterno” vengono spesso “sentite” da recettori, trasmesse al nucleo attraverso “vie di trasduzione del segnale” dove il “messaggio” attiva i geni appropriati per la risposta cellulare. In questa catena di eventi, i fattori di trascrizione sono il passaggio finale delle vie di trasduzione del segnale dall'esterno all'interno. Vediamo ora tre esempi di attivazione della trascrizione genica in risposta a stimoli esterni per cercare di comprendere come si realizza la regolazione trascrizionale di specifici geni.

Recettori degli ormoni steroidei

Questa famiglia di attivatori trascrizionali, già presentata nello schema di **figura 11.21**, ha una sequenza amminoacidica molto conservata, con il domino di legame al DNA che appartiene al motivo a dita di zinco. Hanno, inoltre, un dominio di attivazione e un dominio di legame all'ormone specifico, un dominio di dimerizzazione e un segnale di localizzazione nucleare. I ligandi che attivano questi fattori sono piccoli, passano facilmente la membrana cellulare in quanto liposolubili e si legano direttamente agli attivatori corrispondenti. Gli ormoni steroidei, l'ormone della tiroide, la vitamina D e i retinoidi, si legano a membri specifici di questa famiglia di attivatori permettendo il loro legame ai siti specifici sul DNA, a monte dei geni che sono sotto il loro controllo. In particolare si legano a sequenze del tipo 5'AGAACA o 5'AGGTCA, che permettono al recettore di legarsi come dimero.

Dato l'elevato grado di conservazione della loro sequenza amminoacidica, alcuni di questi attivatori possono legarsi come eterodimeri con membri della stessa famiglia, ma non è escluso che si possano formare eterodimeri anche con altri tipi di attivatori.

Questa capacità di formare eterodimeri è importante per la possibilità di accoppiare la risposta agli ormoni con altre vie del segnale. In assenza del ligando, l'attivatore-recettore si trova nel citoplasma legato alla proteina Hsp90 (coinvolta nella risposta allo shock termico), la quale maschera il suo segnale di localizzazione nucleare. Dopo il legame con l'ormone, la proteina cambia conformazione e Hsp90 si dissocia permettendo al recettore di entrare nel nucleo, dove si lega al DNA e attiva il gene bersaglio (**fig. 11.30A**).

Risposta all'AMP ciclico e fattore CREB

Come abbiamo appena visto, l'espressione di alcuni geni può essere legata alla risposta che molecole segnale inducono interagendo con recettori esposti sulla superficie della membrana cellulare. Il ligando induce dei cambiamenti strutturali nel recettore, il quale innesca una serie di eventi che, alla fine, portano a una variazione dell'attività di trascrizione all'interno del nucleo. Una delle modificazioni che avvengono in seguito all'attivazione di una via di trasduzione del segnale è la fosforilazione di un attivatore che diventa attivo dopo questa modificazione. Come mostrato in **figura 11.30B**, un recettore a sette eliche transmembrana che

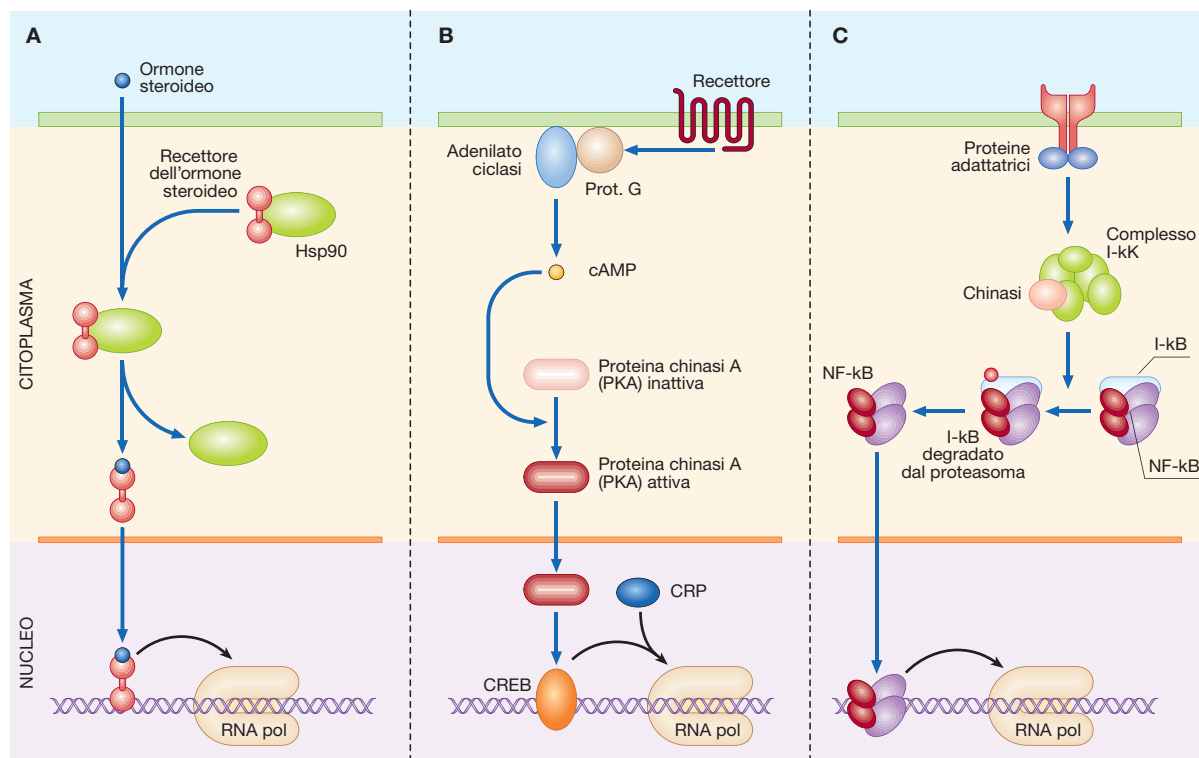


Figura 11.30 Tre esempi di attivazione della trascrizione in risposta a stimoli esterni tramite il sistema di trasduzione del segnale. (A) Gli ormoni steroidei diffondono all'interno della membrana cellulare e raggiungono un recettore che si trova nel citoplasma legato alla proteina Hsp90. Il legame con l'ormone provoca un cambio conformazionale e il recettore si stacca dal complesso ed è trasportato nel nucleo, dove si lega al DNA e promuove la trascrizione dei geni specifici per la risposta all'ormone. (B) Un ligando che si lega a un recettore a sette domini di membrana innescava una serie di segnali che portano all'attivazione di geni specifici. Il le-

game del ligando attiva una proteina G che, a sua volta, induce la produzione di AMP ciclico. Questa molecola attiva una proteina chinasi che migra nel nucleo, dove fosforila il fattore trascrizionale CREB (*cAMP Response Element Binding Protein*). CRP = *Creb Related Protein*. (C) Fattori di trascrizione costitutivamente attivi vengono trattenuti nel citoplasma fino a che non si attiva una via di segnalazione. In questo caso l'attivatore NF-kB è sequestrato nel citoplasma da un complesso che contiene I-kB. La fosforilazione di I-kB porta alla sua degradazione e alla liberazione di NF-kB, che può migrare nel nucleo e attivare la trascrizione dei geni bersaglio.

si trova sulla superficie di una cellula può essere attivato da uno stimolo. Questo stimolo induce un cambiamento conformazionale nel recettore, l'attivazione di una proteina G che, a sua volta, attiva l'enzima adenilato ciclasi che produce AMP ciclico. Questa molecola è un tipico "secondo messaggero" che attiva una proteina chinasi che, migrando nel nucleo, fosforila il fattore trascrizionale CREB (*cAMP Response Element Binding Protein*) richiesto per l'attivazione di vari geni bersaglio. CREB si lega a siti sul DNA chiamati CRE (*cAMP Response Element*). CREB si lega al DNA come dimero, facendo parte della famiglia dei fattori trascrizionali con il motivo a cerniera di leucine. I geni regolati da CREB includono: *c-fos*, la neurotropina BDNF (*Brain-Derived Neurotrophic Factor*) e molti neuropeptidi come la somatostatina, l'encefalina e l'ormone di rilascio della corticotropina. CREB è strettamente correlato, sia per struttura che per funzione, a CREM (*cAMP Response Element Modulator*) e ad ATF-1 (*Activating Transcription Factor-1*). CREB ha un ruolo molto importante nella plasticità dei neuroni e nella formazione della memoria a lungo termine.

Fattore trascrizionale NF-kB

Le proteine della famiglia NF-kB sono coinvolte nella trascrizione di geni espressi in vari tipi di cellule; abbiamo già visto che una di queste proteine è presente nell'enhanceosoma dell'interferone β . Come mostrato in **figura 11.30C**, NF-kB

rimane sequestrato nel citoplasma tramite il legame con l'inibitore noto come I-kB finché una via di trasduzione del segnale non porta alla fosforilazione dell'inibitore e alla sua degradazione da parte del proteosoma (*vedi* par. 19.1). Dopo il distacco dall'inibitore, NF-kB è libero di migrare nel nucleo e di attivare la trascrizione dei geni bersaglio.

Lecture di approfondimento

Pol I

McStay, B., Grummt, I. (2008), «The epigenetics of rRNA genes: from molecular to chromosome biology», *Annual Review of Cell and Developmental Biology*, **24**, pp. 131-157.

Russell, J., Zomerdijs, J.C. (2005), «RNA-polymerase-I-directed rDNA transcription, life and works», *Trends in Biochemical Sciences*, **30**(2), pp. 87-96.

Pol II

Biggin, M.D. (2011), «Animal transcription networks as highly connected, quantitative continua», *Developmental Cell*, **21**, pp. 611-626.

Buratowski, S. (2008), «Transcription. Gene expression--where to start?», *Science*, **322**(5909), pp. 1804-1805.

Malik, S. e Roeder, R.G. (2005), «Dynamic regulation of pol II transcription by the mammalian Mediator complex», *Trends in Biochemical Sciences*, **30**(5), pp. 256-263.

Juven-Gershon, L. et al. (2008), «The RNA polymerase II core promoter - The gateway to transcription», *Current Opinion in Cell Biology*, **20**(3), pp. 253-259.

Kornberg, L. (2007), «The molecular basis of eukaryotic transcription», *Proc. Natl. Acad. Sci. USA*, **104**(32), pp. 12955-12961.

Ptashne, M. (2008), «Transcription: a mechanism for short-term memory», *Current Biology*, **8**, pp. 18(1).

Pol III

Dieci, G. et al. (2007), «The expanding RNA polymerase III transcriptome», *Trends in Genetics*, **23**(12), pp. 614-622.

Geiduschek, E.P. (2010), «An Introduction to Transcription and Gene Regulation», *Journal of Biological Chemistry*, **285**(34), pp. 25885-25892.

Geiduschek, E.P., Kassavetis, G.A. (2006), «Transcription: adjusting to adversity by regulating RNA polymerase», *Current Biology*, **16**(19), pp. R849-851.

E

PROCESSAMENTO E MATURAZIONE DELL'RNA

La trascrizione, soprattutto negli eucarioti in cui è confinata nel nucleo, non produce generalmente RNA funzionali, ma molecole di precursore che potranno esercitare la loro attività solo dopo aver subito una serie di processi di maturazione che includono:

- tagli nucleolitici e modificazioni chimiche dell'RNA, descritti nel capitolo 12;
- modificazioni delle sequenze nucleotidiche mediante lo "splicing" e l'"editing" dell'RNA, descritte nel capitolo 13.

Per esempio, l'azione di esonucleasi ed endonucleasi è fondamentale per la risoluzione di trascritti policistronici in trascritti monocistronici, come nel caso dei precursori degli RNA ribosomali, e per la formazione di tRNA maturi in seguito alla rimozione di tratti oligonucleotidici in eccedenza al 5' e al 3'. Le modificazioni chimiche degli RNA, come l'aggiunta di gruppi metilici, la formazione di basi insolite come la pseudouridina, l'aggiunta del 5' cap e della coda di poli(A), rispettivamente alle estremità 5' e 3' dell'mRNA, rappresentano un altro importante processo di maturazione per la formazione di RNA funzionali in grado di concorrere al processo della traduzione.

La produzione di RNA messaggeri dotati dell'informazione corretta per guidare la traduzione richiede, infine, la rimozione degli introni attraverso lo splicing e talvolta anche la modificazione post-trascrizionale della loro porzione codificante attraverso il processo di RNA editing. Tutti questi processi verranno trattati più dettagliatamente nei prossimi capitoli.

Negli eucarioti i processi sopra menzionati avvengono nel nucleo in modo da evitare che l'RNA precursore o parzialmente maturato possa essere utilizzato impropriamente dall'apparato di traduzione localizzato nel citoplasma. Solo gli RNA correttamente maturati vengono traslocati dal nucleo al citoplasma attraverso un appropriato sistema di trasporto attivo.

I processi di maturazione, che sono fondamentali per rendere funzionalmente attivi gli RNA prodotti, costituiscono un ulteriore livello di regolazione dell'espressione genica e sono anche responsabili dell'espansione del potenziale di espressione di un genoma sia a livello del trascrittoma che del proteoma. Per esempio, come vedremo più avanti, lo splicing alternativo e l'editing consentono la produzione di trascritti e proteine alternative a partire da uno stesso gene, anche in modo tessuto-specifico.

12

Maturazione dell'RNA: tagli nucleolitici e modificazioni chimiche

12.1 Processamento degli rRNA nei procarioti e negli eucarioti

12.2 Processamento dei tRNA nei procarioti e negli eucarioti

12.3 Ribozimi autocatalitici a “testa di martello”

12.4 Modificazioni chimiche delle basi e del ribosio

12.5 Aggiunta dei “cap” agli mRNA eucariotici

12.6 Poliadenilazione e terminazione della trascrizione degli mRNA eucariotici

12.1 Processamento degli rRNA nei procarioti e negli eucarioti

I geni codificanti per gli rRNA ribosomali (rRNA) sia dei procarioti che degli eucarioti vengono trascritti sotto forma di grandi molecole di RNA precursore che poi daranno origine alle molecole di rRNA mature in seguito all'escissione delle porzioni in eccesso. Questo meccanismo denominato **processamento** (*processing*) è diverso dallo splicing perché gli RNA maturi vengono generati da tagli endonucleolitici, eventualmente seguiti da un'azione di rifinitura con esonucleasi 3'→5' o 5'→3', ma non si ha la concatenazione dei frammenti prodotti come nello splicing. Descriveremo ora questo processo prima nei procarioti e poi negli eucarioti.

Per illustrare il processo di maturazione nei procarioti ci riferiremo a quanto si osserva in *Escherichia coli*. I geni per gli rRNA di *E. coli* sono presenti in sette copie, organizzati in altrettanti operoni ciascuno dei quali contiene una copia di ciascun tipo di rRNA (rRNA 5S, rRNA 16S e rRNA 23S). In **figura 12.1A** è schematizzata la struttura di uno degli operoni (*rrnD*), che oltre ad avere una copia di ciascun rRNA contiene anche tre geni per tRNA. La trascrizione dell'operone genera un RNA precursore 30S che poi subirà un processing che richiede l'azione della RNasi III. Questo enzima effettua tagli endonucleolitici in due regioni complementari che fiancheggiano sia l'rRNA 16S che 23S (**fig. 12.1B**). Un'altra ribonucleasi, RNasi E, è responsabile della rimozione dell'rRNA 5S dalla molecola di precursore. I tRNA saranno invece maturati con il meccanismo descritto nel paragrafo 12.2. Una descrizione schematica dei vari tipi di nucleasi impegnate nel processamento degli RNA è riportata nella **finestra 12.1**.

Negli eucarioti i geni che codificano per il precursore degli rRNA (pre-rRNA) sono ripetuti in tandem e localizzati su diversi cromosomi che contengono centinaia di copie di ciascun gene (**fig. 12.2**). Ogni gene per pre-rRNA è separato dal successivo da una regione spaziatrice non trascritta (NTS, *Non Transcribed Spacer*).

La trascrizione dei cluster di rRNA può anche essere osservata al microscopio

FINESTRA 12.1 LE NUCLEASI

In questo capitolo sul processamento dell'RNA, ma anche in vari altri capitoli del libro, incontriamo frequentemente enzimi capaci di tagliare le catene polinucleotidiche rompendo i legami fosfodiesterici tra nucleotidi adiacenti. Tali enzimi, chiamati genericamente **nucleasi**, sono molto numerosi nella cellula e molto eterogenei per grandezza, struttura, meccanismo d'azione e per il substrato su cui agiscono. Ne diamo qui una breve e schematica descrizione e la **figura 12F1.1** schematizza graficamente l'attività dei vari tipi di nucleasi.

Le nucleasi comprendono le DNasi (deossiribonucleasi) specifiche per il DNA, le RNasi (ribonucleasi) specifiche per l'RNA e alcuni enzimi in grado di tagliare sia il DNA che l'RNA per i quali manteniamo il nome generico di nucleasi.

Alcune nucleasi, sia DNasi che RNasi, tagliano la molecola di acido nucleico in posizione interna e vengono chiamate **endonucleasi**, altre rimuovono i nucleotidi a partire dalle estremità e vengono chiamate **esonucleasi**.

Tra le endonucleasi distinguiamo quelle che riconoscono e tagliano la doppia elica (I) e quelle che riconoscono e tagliano il singolo filamento di acido nucleico (II). Esempi delle prime sono gli enzimi di restrizione (vedi cap. 20), la nucleasi micrococcica e le DNasi I e II (vedi par. 4.8), e l'RNasi III. Esempi delle seconde sono la tRNA 3' endonucleasi e l'RNasi P, la nucleasi S1 (vedi cap. 20) e i "cleaving factor" CFI e CFII (vedi par. 12.6).

A causa della polarità 5'-3' delle catene polinucleotidiche, le esonucleasi vengono distinte in **esonucleasi 5'-3' (III)** ed **esonucleasi 3'-5' (IV)** a seconda di quale estremità della molecola attaccano. Esempi delle prime sono la Xrn1p (vedi cap. 16) e il dominio 5'-3' exo della DNA polimerasi I di *E. coli* (vedi cap. 6); esempi delle seconde sono la PNPasi di *E. coli* e l'esosoma del lievito (vedi par. 16.2) e il dominio "correttore di bozze" delle DNA polimerasi (vedi cap. 6).

Per quanto riguarda il meccanismo d'azione, alcune esonucleasi vengono dette **processive**, perché, una volta che l'enzima si è attaccato all'estremità della molecola di acido nucleico e ha rimosso il primo nucleotide, poi prosegue nella rimozione del secondo, del terzo ecc. (V); sono, per esempio, processive la exoI di *E. coli* e l'esonucleasi di lambda. Invece le esonucleasi di **tipo non processivo**, dopo avere rimosso il nucleotide all'estremità, si staccano dal polinucleotide substrato che viene aggredito da un'altra molecola di esonucleasi che rimuove il secondo nucleotide, per poi a sua volta distaccarsi ecc. (VI); ne sono esempio l'esonucleasi III di *E. coli* e l'esonucleasi T7. La maggior parte delle nucleasi, endo ed exo, tagliano a monte del gruppo fosfato lasciando questo gruppo al 5' del nucleotide seguente (VII); alcune, al contrario, tagliano a valle del gruppo

fosfato lasciando questo al 3' del nucleotide a monte (VIII) come avviene, per esempio, durante lo splicing degli introni contenuti nei tRNA (vedi cap. 13).

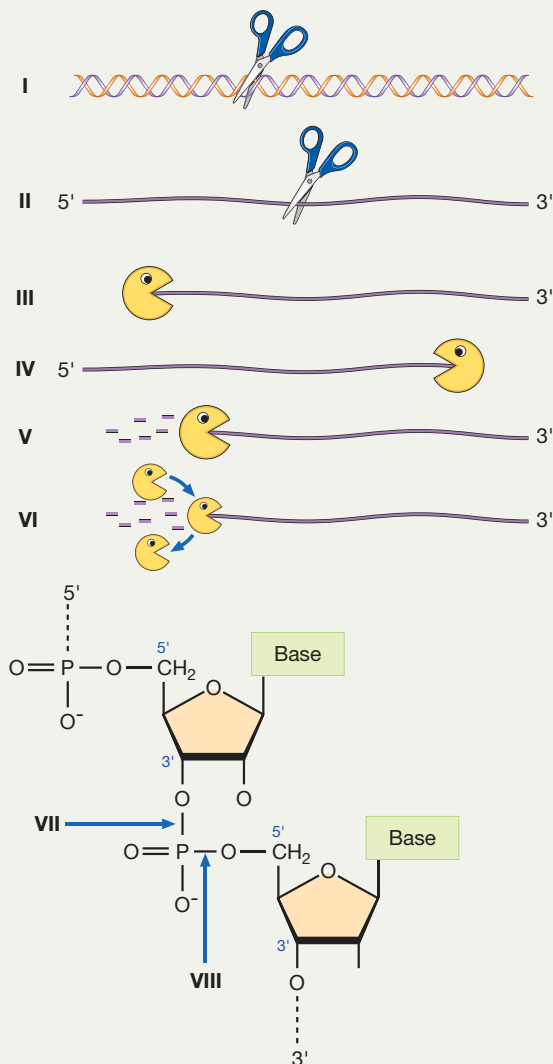


Figura 12F1.1 Rappresentazione schematica delle attività delle diverse nucleasi. (I) Endonucleasi a doppia elica. (II) Endonucleasi a singolo filamento. (III) Esonucleasi 5'-3'. (IV) Esonucleasi 3'-5'. (V) Esonucleasi processiva. (VI) Esonucleasi non processiva. (VII) La maggior parte delle nucleasi taglia a monte del gruppo fosfato, lasciando questo legato al 5' del ribosio successivo. (VIII) Alcune nucleasi tagliano a valle del gruppo fosfato, lasciando questo legato al 3' del ribosio precedente.

elettronico (**fig. 12.3**), dove si distinguono gli NTS dalle regioni trascritte che appaiono nella forma di "alberi di Natale" in cui i rami più lunghi alla base rappresentano le molecole di RNA precursore più estese.

L'RNA polimerasi I dei mammiferi genera una molecola di pre-rRNA 45S che contiene al suo interno i geni per gli rRNA 28S, 18S e 5.8S, separati tra loro da regioni spaziatrici denominate ITS (*Internal Transcribed Spacer*) (**fig. 12.2**). I geni per

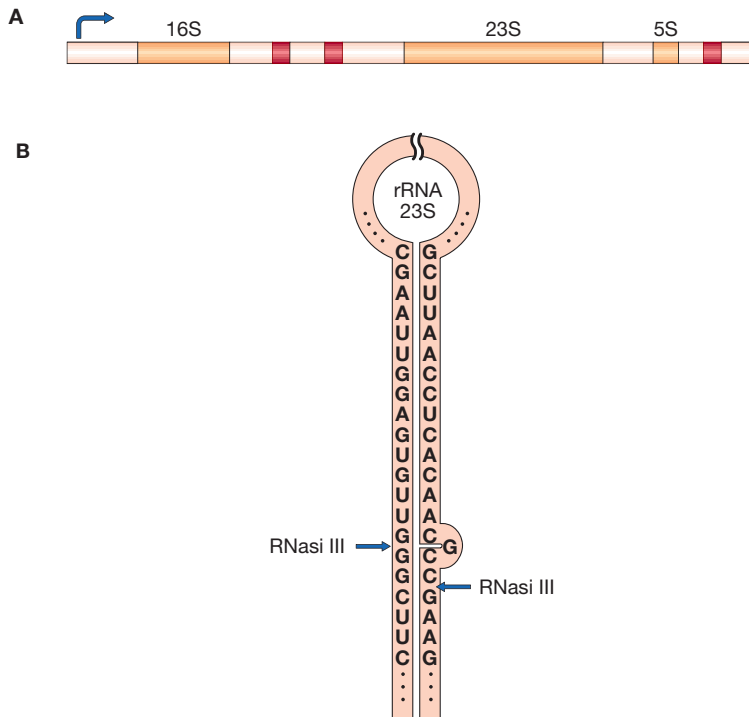


Figura 12.1 Struttura di un operone *rrn* di *Escherichia coli* contenente i geni ribosomali 16S, 23S e 5S. (A) Rappresentazione schematica della regione di DNA contenente l'operone *rrnD*, uno dei sette operoni *rrn* di *E. coli*; la freccia blu parte dal promotore e indica la direzione della trascrizione; in rosso sono indicati i tre geni per tRNA contenuti in questo operone. (B) Regione a doppia elica che si forma per appaiamento delle regioni complementari fiancheggianti la sequenza dell'rRNA 23S e che funge da substrato per il taglio endonucleolitico catalizzato dall'RNasi III.

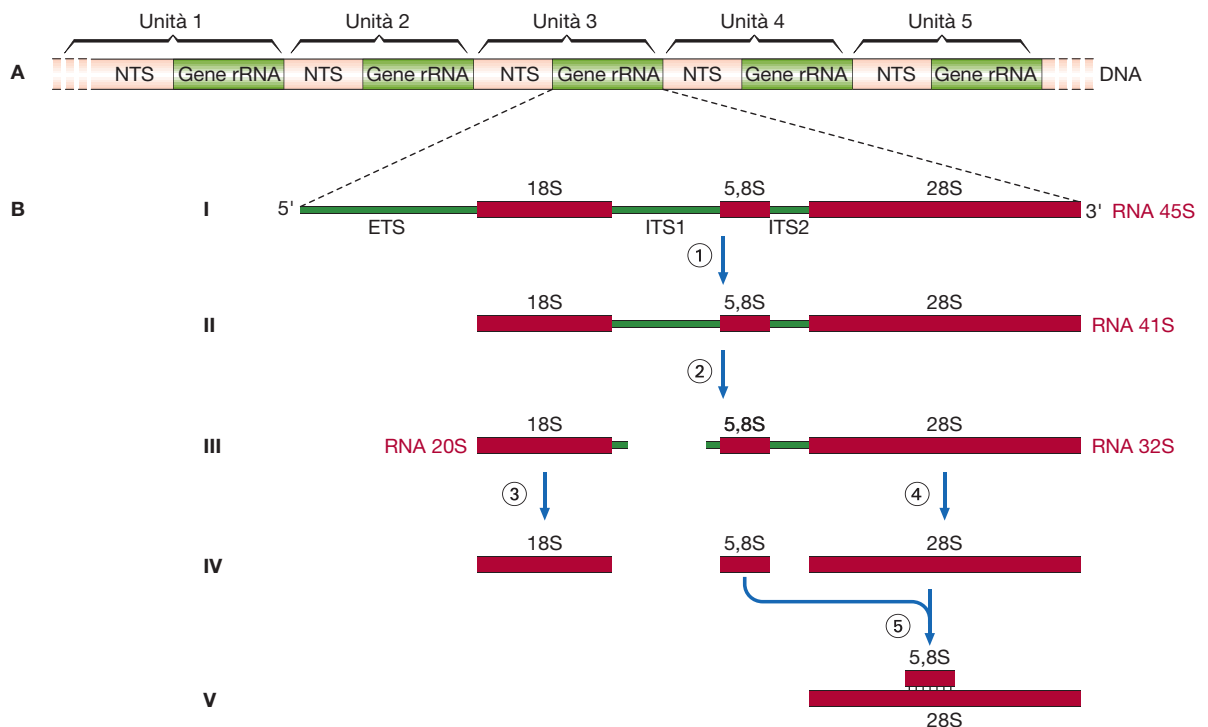
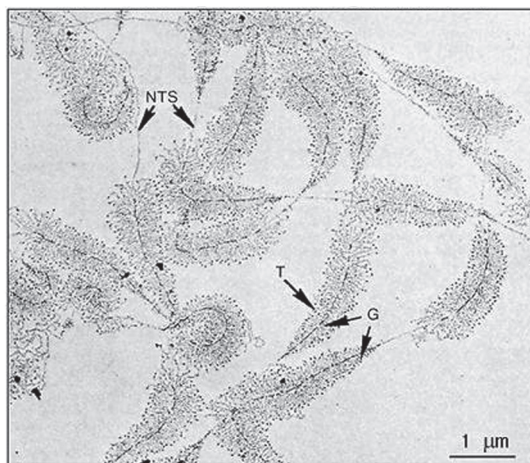


Figura 12.2 Struttura dei geni per il precursore dell'rRNA nelle cellule di mammiferi. I geni per gli rRNA sono ripetuti in tandem in una o più localizzazioni nel genoma. (A) Ciascuna unità ripetitiva è costituita da una regione non trascritta (NTS, *Non Transcribed Spacer*) e da una regione trascritta che comprende una ETS (*External Transcribed Sequence*), il gene per il 18S rRNA, l'ITS1 (*Internal Transcribed Spacer 1*), il gene per l'rRNA 5,8S, l'ITS2 e il gene per l'rRNA 28S. (B) La trascrizione risulta nella sintesi di un lungo precursore dell'rRNA, chiamato RNA 45S (I), la cui maturazione include le seguenti fasi. (1) Rimozione dell'estremità 5' e formazione del precursore 41S (II). (2) Il precursore 41S viene scisso in due parti, il precursore 20S dell'rRNA 18S e il precursore 32S degli rRNA 5,8S e 28S (III). (3) Viene rimossa l'estremità 3' del precursore 20S formando l'rRNA 18S maturo, mentre (4) il taglio del precursore 32S libera gli rRNA maturi 5,8S e 28S (IV), che si associano (5) per complementarità delle basi (V).

precursore dell'rRNA, chiamato RNA 45S (I), la cui maturazione include le seguenti fasi. (1) Rimozione dell'estremità 5' e formazione del precursore 41S (II). (2) Il precursore 41S viene scisso in due parti, il precursore 20S dell'rRNA 18S e il precursore 32S degli rRNA 5,8S e 28S (III). (3) Viene rimossa l'estremità 3' del precursore 20S formando l'rRNA 18S maturo, mentre (4) il taglio del precursore 32S libera gli rRNA maturi 5,8S e 28S (IV), che si associano (5) per complementarità delle basi (V).

Figura 12.3 Immagine al microscopio elettronico del cluster dei geni (G) che codificano per i precursori dell'RNA ribosomale di anfibio. Le molecole di RNA del precursore 45S (T) attivamente trascritte appaiono nella caratteristica struttura ad "albero di Natale". Alla base di ogni trascritto in fase di sintesi vi è una molecola di RNA polimerasi I (non visibile). I geni sono separati da regioni spaziatrici non trascritte (NTS). [Fonte: Miller, O.L., Jr., Beatty, B.R., Hamkalo, B.A. e Thomas, C.A. Jr. (1970), «Electron Microscopic Visualization of Transcription», *Cold Spring Harb. Symp. Quant. Biol.*, 35, pp. 505-512, doi:10.1101/SQB.1970.035.01.066.]



l'rRNA 5S, anch'essi presenti in centinaia di copie disposte in tandem, hanno una localizzazione differente nel genoma e vengono trascritti dall'RNA polimerasi III.

La sintesi e il processing degli rRNA, come anche l'assemblaggio dei ribosomi che descriveremo nel paragrafo 12.4, hanno luogo nel nucleolo (**finestra 12.2**). In **figura 12.2** sono illustrate le fasi del processing, che consiste in una serie di tagli endonucleolitici in successione fino alla formazione delle molecole degli rRNA maturi.

Il processing e la maturazione degli rRNA richiedono l'intervento di una classe di piccoli RNA nucleolari denominati **snoRNA** (*small nucleolar RNA*) sotto forma di particelle ribonucleoproteiche (snoRNP) (*vedi par. 12.4*). Per esempio, gli snoRNA E1 ed E2 partecipano alla maturazione dell'rRNA 18S, mentre lo snoRNA E3 è coinvolto nella maturazione dell'rRNA 5,8S.

12.2 Processamento dei tRNA nei procarioti e negli eucarioti

Sia nei procarioti che negli eucarioti, i tRNA vengono trascritti sotto forma di precursori che dovranno subire un processo di maturazione per generare i tRNA funzionali.

Nei procarioti l'RNA precursore può contenere una o più molecole di tRNA o, come abbiamo già visto (**fig. 12.1**), sia molecole di tRNA che di rRNA. La prima fase del processo di maturazione o processamento comporterà, dunque, la risoluzione dell'RNA precursore policistronico in una serie di molecole di RNA monocistroniche. L'enzima che entra in gioco in questa fase è l'RNasi III, un'endonucleasi che genera le molecole di pre-tRNA. Questa fase non è necessaria negli eucarioti, in cui gli RNA precursori sono generalmente monocistronici.

I pre-tRNA procariotici ed eucariotici contengono estremità più estese sia al 5' che al 3' che dovranno essere rimosse dall'azione di endonucleasi ed esonucleasi per generare le molecole di tRNA mature e funzionali.

L'estremità 5' del tRNA maturo viene generata da un taglio endonucleolitico catalizzato dall'enzima RNasi P, una ribonucleoproteina la cui componente primaria è l'RNA M1 che esercita l'attività catalitica (**fig. 12.4**).

La formazione dell'estremità 3' è un processo molto più complesso e utilizza meccanismi differenti in diversi organismi. La maturazione dei pre-tRNA batterici richiede un taglio endonucleolitico a valle della tripletta CCA, che diversamente dai tRNA eucariotici è già presente nel precursore, seguita dall'azione di una varietà di esonucleasi che includono RNasi II, BN, D, PH, T e PNPasi, che completano il processo (**fig. 12.4A**).

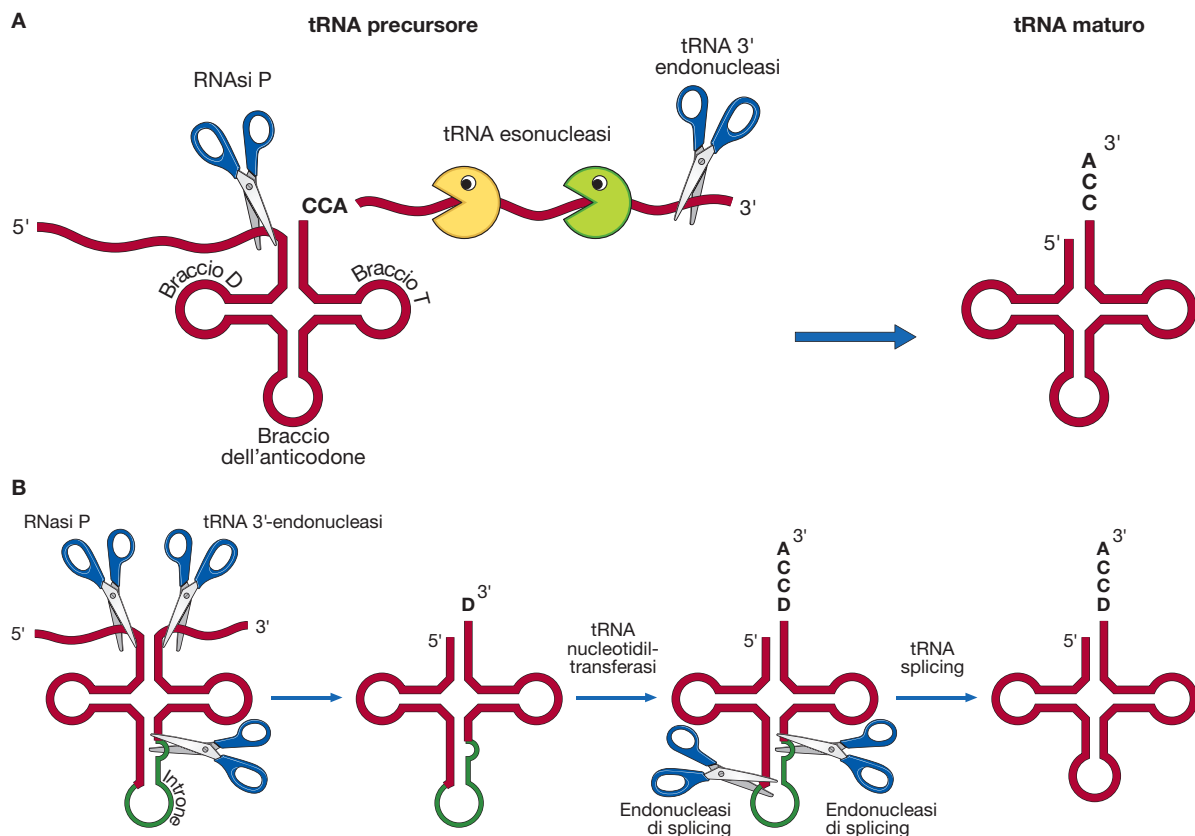


Figura 12.4 Maturazione dei tRNA. (A) La maturazione dei tRNA in *Escherichia coli*. (1) Una tRNA 3' endonucleasi effettua un taglio nella regione 3' del precursore; (2) varie tRNA esonucleasi rimuovono i nucleotidi in eccesso fino all'estremità 3' del tRNA maturo; (3) un taglio effettuato dall'RNasi P genera l'estremità 5' del tRNA maturo. (B) La maturazione dei tRNA eucariotici. (1) Una tRNA 3' endonucleasi effettua un taglio immediatamente a valle del nu-

cleotide discriminante (D), importante per il riconoscimento specifico da parte della aminoacil-tRNA sintetasi (vedi par. 14.2); il taglio da parte dell'RNasi P genera l'estremità 5' del tRNA; (2) la tRNA nucleotidil transferasi aggiunge la tripletta CCA al 3'; (3) se il precursore del tRNA contiene un introne, questo viene rimosso da una specifica endonucleasi (vedi fig. 13.14).

I pre-tRNA eucariotici hanno estensioni al 3' di 5-15 nucleotidi e generalmente non contengono la tripletta terminale CCA. La reazione di processing al 3' è diversa da organismo a organismo, ma generalmente comporta un taglio endonucleolitico all'estremità 3' del tRNA, subito a valle del nucleotide discriminante (fig. 12.4B) da parte di una tRNasi o RNasi Z, seguita dall'aggiunta del trinucleotide CCA da parte della tRNA nucleotidil transferasi.

Anche nei mitocondri e nei cloroplasti i tRNA vengono sintetizzati sotto forma di precursori, generalmente privi della tripletta CCA terminale, che vengono poi maturati secondo un meccanismo simile a quello illustrato per gli eucarioti e che prevede tagli endonucleolitici da parte dell'RNasi P al 5', l'azione di un'altra endonucleasi specifica al 3', seguita poi dall'aggiunta della tripletta CCA. Dato che gli RNA precursori di alcuni tRNA possono contenere introni, questi verranno rimossi per l'azione di una specifica endonucleasi deputata allo splicing (vedi par. 13.1).

La maturazione dei tRNA viene poi completata da una serie di modificazioni dei nucleotidi operate da fattori enzimatici specifici, come descriveremo nel paragrafo 12.4.

12.3 Ribozimi autocatalitici a “testa di martello”

I **ribozimi a testa di martello** (*hammerhead*) sono costituiti da piccole molecole di RNA (200-400 nt) in grado di effettuare un taglio autocatalitico che porta alla

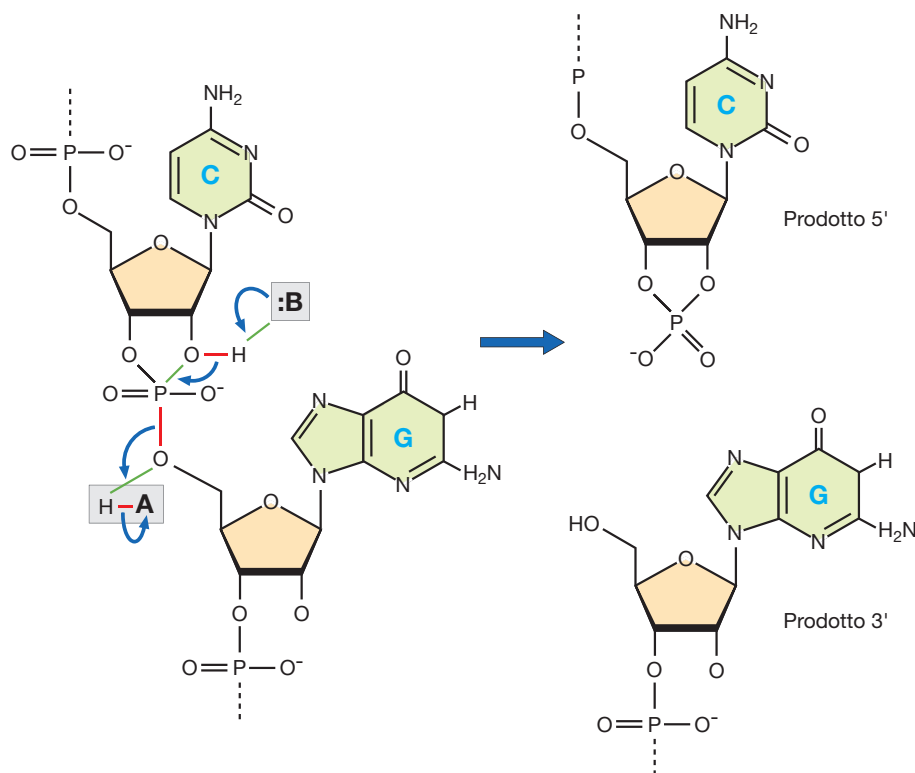


Figura 12.5 Meccanismo catalitico degli enzimi a testa di martello. L'azione idrolitica, che comporta la formazione di uno stato di transizione pentacoordinato, inizia con la deprotonazione del gruppo ossidrilico al 2', favorita da una base debole (:B) cui segue l'idrolisi del legame fosfoesterico favorita da un acido debole (H-A). Il prodotto di reazione al 5' è un fosfato 2'-3' ciclico. I

legami in fase di rottura sono mostrati in rosso, mentre quelli in formazione in verde. Le frecce curve blu rappresentano il movimento degli elettroni. Alcune evidenze suggeriscono che gli ioni magnesio bivalenti in equilibrio tra la forma esaidrata e quella di idrossido possano fungere sia da agente acido che basico, rispettivamente.

formazione di un fosfato ciclico 2'-3' all'estremità 3' e di un ossidrilico all'estremità 5' (**fig. 12.5**). L'attività funzionale dei ribozimi hammerhead consiste nel processamento dei piccoli genomi a RNA di alcuni viroidi patogeni delle piante, richiesto per la loro replicazione, che avviene secondo il modello del cerchio rotante (*rolling circle*).

I viroidi sono costituiti da RNA circolare a singolo filamento che non codifica per alcuna proteina. Essi esercitano la loro attività patogena attraverso il meccanismo di silenziamento genico mediato dall'interazione complementare con un mRNA bersaglio. Gli intermedi a doppio filamento della replicazione dei viroidi possono, infatti, fungere da substrato per enzimi della classe DICER ed essere processati per formare siRNA che poi potranno esercitare la loro attività di silenziamento una volta caricati sul complesso RISC (*RNA-Induced Silencing Complex*).

I ribozimi hammerhead sono caratterizzati da un core di 13-15 nucleotidi conservati da cui si irradiano tre forchine a doppia elica di lunghezza variabile (**fig. 12.6**).

Utilizzando le tecniche dell'ingegneria genetica è possibile realizzare ribozimi hammerhead artificiali costituiti da due distinte molecole di RNA, una contenente il sito di taglio che funge da substrato e l'altra che esercita l'attività enzimatica. I ribozimi ricombinanti così realizzati possono avere importanti applicazioni biotecnologiche, soprattutto per effettuare silenziamenti genici mirati per combattere malattie virali come l'AIDS e l'epatite B, e altre importanti patologie come il cancro, il diabete, l'artrite reumatoide e la sclerosi laterale amiotrofica. In **figura 12.7** è mostrata la struttura di un ribozima hammerhead progettato per silenziare il gene fushi-tarazu (*ftz*), coinvolto nella segmentazione del blastoderma durante lo sviluppo di *Drosophila*.

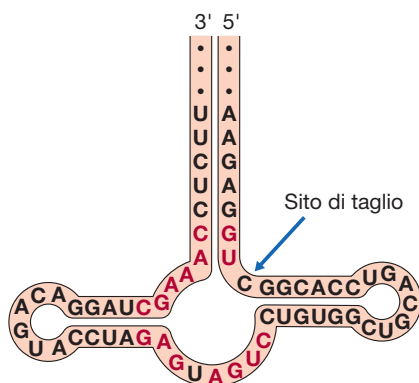


Figura 12.6 Struttura consenso di un ribozima hammerhead naturale. È qui mostrata la sequenza e la struttura della testa di martello del virusoide vegetale CChMVd. La freccia indica il sito del taglio autocatalitico. Le basi conservate nelle varie strutture a testa di martello sono segnate in rosso.

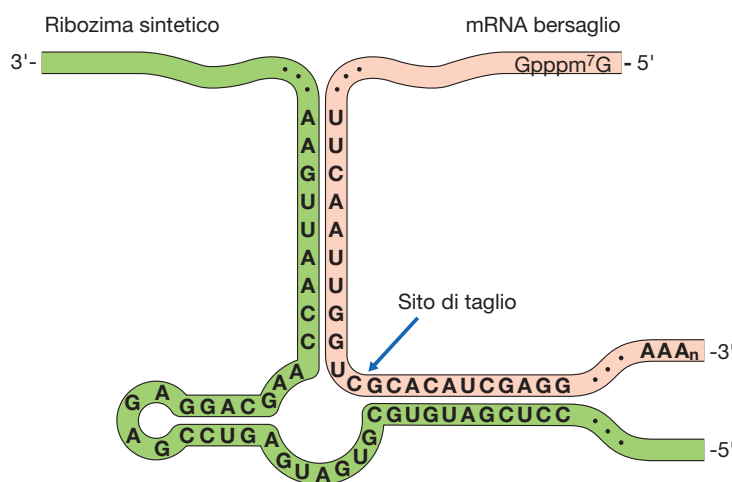


Figura 12.7 Struttura di un ribozima hammerhead sintetico legato all'mRNA bersaglio. Interazione complementare tra l'RNA messaggero espresso dal gene fushi-tarazu di *Drosophila* (in rosa) e un ribozima hammerhead sintetico (in verde) che attiva il taglio endonucleolitico dell'mRNA silenziando così il gene.

12.4 Modificazioni chimiche delle basi e del ribosio

Nel corso del processo di maturazione dell'RNA, numerosi enzimi introducono delle modificazioni chimiche sia delle basi azotate, a livello di atomi di carbonio (ad es. posizione 5 della citosina) e di azoto (ad es. posizione 7 della guanosina), che del ribosio, normalmente a livello del gruppo ossidrilico al 2' (ad es. formazione di 2'-idrossimetile). Nelle basi azotate, oltre a quelle negli esempi sopra citati, altre posizioni frequentemente modificate sono i gruppi amminici in posizione 6 dell'adenina e in posizione 2 della guanina, come pure nella posizione 3 della citosina. Tali modificazioni vengono in genere effettuate direttamente sull'RNA già formato. In alcuni casi, basi modificate possono essere introdotte successivamente; per esempio, il cap al 5' degli mRNA eucariotici (par. 12.5) è costituito da una 7-metil guanosina introdotta post-trascrizionalmente da un'attività nucleotidil transferasica. La localizzazione, abbondanza e distribuzione dei vari tipi di modificazioni è molto variabile in diverse molecole di RNA, in diversi organismi e compartimenti cellulari.

I ribonucleosidi modificati sono molto abbondanti nelle molecole di tRNA e rRNA, ma sono anche presenti in RNA spliceosomiali (snRNA), nucleolari (snoRNA) e piccoli RNA regolatori come siRNA, miRNA e piRNA. Sono state descritte oltre 100 modificazioni chimiche dei ribonucleosidi, la maggior parte delle quali (75% circa) si osserva nei tRNA. Tali modificazioni richiedono l'intervento di numerosi enzimi che agiscono in vie metaboliche anche complesse.

Nonostante il ruolo fisiologico dei ribonucleosidi modificati sia ancora largamente sconosciuto, è evidente che questi hanno un ruolo funzionale importante e che vi sono chiare implicazioni in molte patologie dell'uomo, incluse le malattie autoimmuni, il cancro e le virosi.

Le principali modificazioni chimiche presenti nelle molecole di RNA sono state descritte nel paragrafo 2.1 e mostrate in figura 2.2.

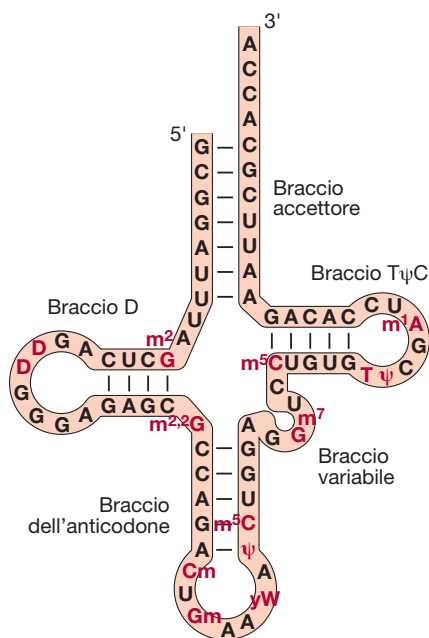
Modificazioni chimiche delle basi nei tRNA procariotici ed eucariotici

Come abbiamo già visto, una delle caratteristiche più rilevanti dei tRNA è l'elevato numero di ribonucleosidi modificati che si osservano in specifiche posizioni, pari generalmente a oltre il 10% del totale in diversi organismi. Un esempio specifico è riportato in **figura 12.8** che mostra la distribuzione dei ribonucleosidi modificati osservati nel tRNA^{Phe} di lievito. Gli enzimi responsabili delle modificazioni chimiche dei ribonucleosidi non sono stati ancora ben caratterizzati e le funzioni della maggior parte delle modificazioni osservate non sono ancora chiare. Tuttavia, gli studi sin qui condotti hanno dimostrato una triplice funzione di queste modificazioni: (1) molte delle modificazioni nell'ansa dell'anticodone regolano l'efficienza della traduzione e la crescita cellulare; (2) le modificazioni nel corpo centrale del tRNA modulano la struttura secondaria e terziaria e la sua stabilità, e l'assenza di alcune modificazioni può attivare specifiche vie degradative; (3) modificazioni specifiche sono fondamentali per determinare l'identità e la specificità dei tRNA.

Modificazioni chimiche degli rRNA eucariotici

Dopo la sintesi del precursore dell'rRNA nel nucleolo e il suo processamento mediante tagli endonucleolitici (par. 12.1), la maturazione degli rRNA prosegue, ancora nel nucleolo (**finestra 12.2**), attraverso una serie di modificazioni post-trascrizionali. Le due principali modificazioni, mostrate in **figura 12.9**, sono la **2'-O metilazione**, cioè la metilazione dei ribosi in posizione 2', e la **pseudouridilazione**, cioè la conversione di uridine in pseudouridine (ψ). Come indicato in **tabella 12.1**, queste modificazioni sono molto numerose, e riguardano specifiche posizioni nella sequenza degli rRNA.

Figura 12.8 I tRNA contengono nucleotidi insoliti. I nucleotidi non canonici vengono prodotti dall'azione di numerosi enzimi che agiscono post-trascrizionalmente. Qui è mostrato, come esempio, il tRNA per la fenilalanina (tRNA^{Phe}) di lievito. Per i nucleotidi modificati (scritti in rosso) sono usate le abbreviazioni convenzionali come di seguito riportato: D, diidrouridina; ψ , pseudouridina; m²G, N²-metilguanina; m^{2,2}G, N²,N²-dimetilguanina; Cm, 2'-O-metilcitosina; Gm, 2'-O-metilguanina; yW, wybutosina; m⁵C, 5-metilcitosina; m⁷G, 7-metilguanina; m¹A, N¹-metiladenina.



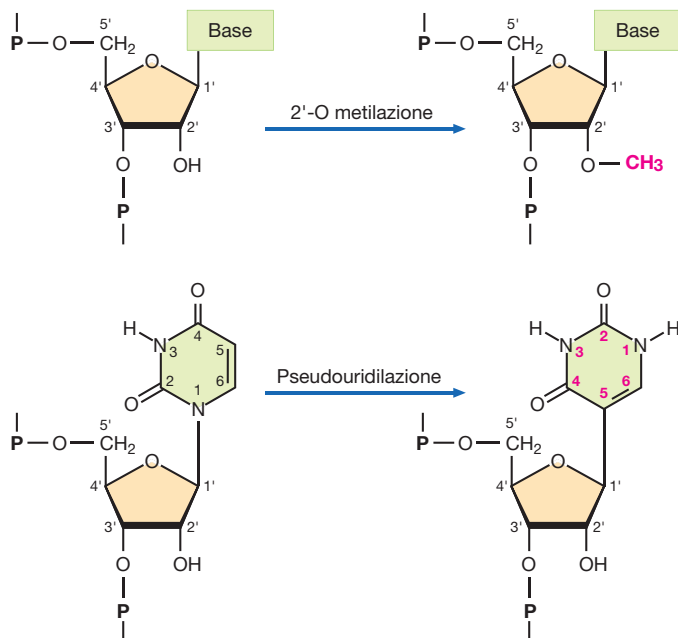


Figura 12.9 Principali modificazioni post-trascrizionali degli rRNA eucariotici. Dopo essere stati sintetizzati nel nucleolo, gli rRNA eucariotici subiscono, nel nucleolo stesso, due tipi di modificazioni post-trascrizionali. (A) 2'-O metilazione: il ribosio di alcuni specifici nucleotidi viene metilato nella posizione 2. (B) Pseudouridilazione: alcune uridine vengono modificate in pseudouridine (Ψ) in cui il ribosio è legato alla posizione 5 dell'uracile anziché alla posizione 1.

FINESTRA 12.2 IL NUCLEOLO

Il nucleolo è una struttura priva di membrana e costituita da proteine e acidi nucleici presente nel nucleo delle cellule eucariotiche, di cui occupa circa il 25% del volume. L'RNA ribosomale è trascritto, maturato e assemblato all'interno del nucleolo, la cui ultrastruttura, che può essere visualizzata con il microscopio elettronico, contiene componenti granulari e fibrillari: la componente fibrillare è costituita da molecole di rRNA associate a proteine, mentre la componente granulare è costituita da subunità ribosomali che sono già state assemblate. Le molecole di 28S, 18S e 5,8S rRNA, derivate dalla molecola di RNA precursore 45S e trascritte all'interno del nucleolo dalla RNA polimerasi I (vedi fig. 12.2), e l'rRNA 5S trascritto dalla RNA polimerasi III, dopo aver subito le modificazioni specifiche mediate dagli snoRNA (vedi fig. 12.10), si assemblano con le proteine ribosomali traslocate nel nucleolo per formare i ribosomi.

La dimensione del nucleolo è correlata al fabbisogno di ribosomi della cellula. Le cellule eucariotiche, di norma, contengono un singolo nucleolo, ma è possibile che ve ne sia più di uno in funzione di specie, tipo di cellula, fase del ciclo cellulare e condizioni fisiologiche. Il nucleolo si forma in corrispondenza di specifici siti cromosomici detti NOR (*Nucleolar Organizer Region*) che contengono i cluster dei geni per il precursore 45S degli rRNA 28S, 18S e 5,8S. Per esempio, nel genoma umano vi sono 10 NOR in corrispondenza dei cluster dei geni per rRNA che sono distribuiti su cinque coppie di cromosomi (13, 14, 15, 21 e 22).

Nel ciclo cellulare il nucleolo viene disgregato all'inizio della mitosi, durante la profase, e si riforma al termine della telofase. Il nucleolo non è solo la sede della formazione dei ribosomi ma è anche coinvolto nella regolazione della progressione

del ciclo cellulare e nella biogenesi della maggior parte delle ribonucleoproteine cellulari.

È emersa recentemente, sulla base di studi condotti in lievito, una possibile connessione tra nucleolo e invecchiamento, che verrebbe accelerato da mutazioni a carico dei geni per rRNA e dalla frammentazione del nucleolo. Questa ipotesi ha ricevuto supporto da studi genetici condotti sulla sindrome di Werner, una patologia ereditaria caratterizzata da invecchiamento precoce.

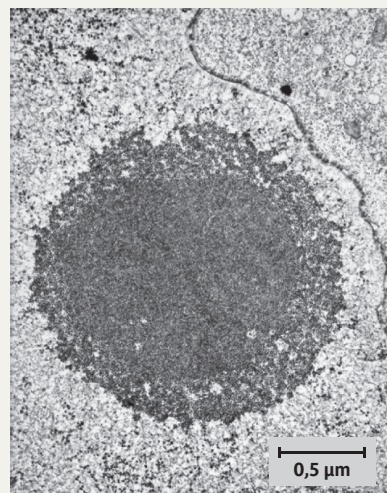


Figura 12F2.1 Il nucleolo del tritone *Notophthalmus viridescens*. Questa fotografia al microscopio elettronico mostra uno dei numerosi nucleoli di un oocita. Sono visibili le tipiche strutture fibrillare (al centro) e granulare (in periferia). (Fonte: Oscar Miller, Dipartimento di Biologia, Università della Virginia.)

Tabella 12.1 Principali modificazioni dei nucleotidi nell'rRNA degli eucarioti: numero approssimato di modificazioni in lievito e nei vertebrati.

Modificazione	Lievito	Vertebrati
Metilazione del 2'-O del ribosio	55	110
Pseudouridilazione	40	100

Tali modificazioni sono guidate, nel nucleolo, da due classi di **snoRNA** (*small nucleolar RNA*) denominati **C/D** e **H/ACA**, dal nome delle brevi sequenze (box) conservate che essi contengono (**fig. 12.10A**). Gli snoRNA presentano una struttura secondaria specifica per ciascuna delle due classi: le due strutture sono schematicamente mostrate nella stessa **figura 12.10A**. Gli snoRNA C/D guidano la 2'-O metilazione, mentre gli snoRNA H/ACA dirigono la modificazione dell'uridina in pseudouridina. In entrambi i casi, i siti che devono subire la modificazione vengono riconosciuti per mezzo di interazioni complementari tra lo snoRNA e l'rRNA bersaglio (*vedi esempio in fig. 12.10B*).

Come era da attendersi, gli snoRNA non agiscono come tali ma assemblati con specifiche proteine in particelle ribonucleoproteiche RNP, che in questo caso prendono il nome di **snoRNP**, in cui le proteine legate esercitano l'attività enzimatica. I numerosi diversi snoRNA C/D sono assemblati con lo stesso gruppo di quattro proteine, e gli snoRNA H/ACA con un diverso gruppo di proteine, come mostrato in **tabella 12.2**. In particolare, una delle proteine che si trova nelle snoRNP C/D è stata chiamata **fibrillarina** in quanto trovata localizzata

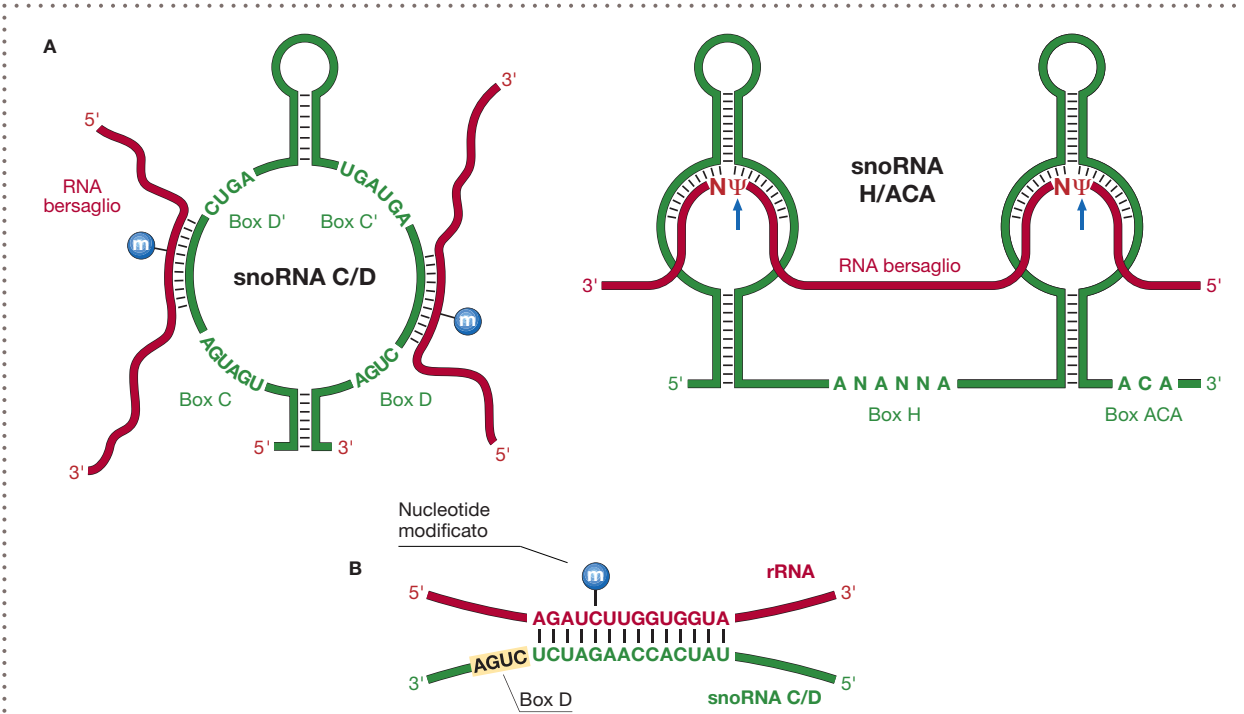


Figura 12.10 Struttura e meccanismo d'azione degli snoRNA C/D e H/ACA. (A) Gli snoRNA C/D e H/ACA (verde) presentano strutture tridimensionali abbastanza conservate e contengono caratteristiche sequenze chiamate box C e box D negli snoRNA C/D (a sinistra) e box H e box ACA negli snoRNA H/ACA (a destra). La presenza negli snoRNA di sequenze complementari ai siti bersaglio dell'rRNA

(rosso) permette il loro riconoscimento per appaiamento di basi e l'identificazione del nucleotide da modificare. Le modificazioni sono indicate in blu: le "m" indicano la metilazione da parte degli snoRNA C/D e le frecce la pseudouridilazione da parte degli snoRNA H/ACA. (B) Dettaglio della regione di complementarità tra uno snoRNA C/D e il corrispondente sito bersaglio sull'rRNA.

Tabella 12.2 Alcune caratteristiche degli snoRNA e delle snoRNP

	snoRNA C/D	snoRNA H/ACA
Funzione	Metilazione del 2'-O del ribosio	Pseudouridilazione
Lunghezza degli snoRNA	70-100 nt	130-140 nt
Proteine associate ai complessi snoRNP	Fibrillarina (Nop1p) NOP56p NOP58p NHPX	Diskarina (pseudouridina sintetasi) Gar1p (NOLA1) NHP2p (NOLA2) NOP10p (NOLA3)

nella porzione fibrillare del nucleolo, mentre una delle proteine delle snoRNP H/ACA è stata chiamata **diskarina** perché sue mutazioni nella specie umana sono alla base della sindrome della discheratosi congenita.

Poiché, come abbiamo detto, gli rRNA ribosomali eucariotici subiscono estese modificazioni con centinaia di siti che vengono metilati o pseudouridilati, esistono nel nucleolo delle cellule eucariotiche almeno altrettanti snoRNA per dirigere correttamente tutte queste modificazioni. Questi stessi tipi di modificazioni possono aver luogo anche su altri RNA bersaglio come i tRNA o gli snRNA spliceosomali (U1, U2, U4, U5, U6), che possono subire entrambi i tipi di modificazione.

12.5 Aggiunta del “cap” agli mRNA eucariotici

Gli mRNA eucariotici vengono generati nel processo trascrizionale (*vedi* cap. 11) all'interno del nucleo a partire da un RNA precursore (anche detto **heterogeneous nuclear RNA**, **hnRNA**) che deve poi subire un processo di maturazione, che consiste in una serie di modificazioni post-trascrizionali, prima di essere traslocato nel citoplasma, dove sarà utilizzato dall'apparato di traduzione per la sintesi della proteina da esso codificata.

La maturazione dell'mRNA precursore avviene nel nucleo e comprende l'aggiunta del cap al 5' (*capping*), di cui parliamo in questo paragrafo, e una serie di altri eventi, che descriveremo nei paragrafi successivi, che include la formazione di una coda di poli(A) all'estremità 3' del trascritto, la rimozione degli introni attraverso il processo di splicing (*vedi* par. 13.1) e alcune modificazioni, note come RNA editing, post-trascrizionali che alterano la sequenza dell'RNA. A seguito dei processi di maturazione sopra elencati, vengono prodotti gli mRNA “maturi” che una volta traslocati nel citoplasma possono essere tradotti in proteine. Alcuni di questi stessi processi di maturazione riguardano non solo gli mRNA ma anche vari tipi di RNA non codificanti.

Il **capping** consiste nell'aggiunta di uno specifico nucleotide modificato con le funzioni di protezione e riconoscimento (da cui la denominazione di *cap*, che si può tradurre cappuccio, tappo, calotta) all'estremità 5' dell'mRNA precursore.

Questa struttura fu scoperta, nel 1974, in seguito alle osservazioni di Bernard Moss e colleghi, che notarono che moltissime specie di mRNA di origine eucariotica o virale risultavano metilate all'estremità 5'. Una serie di esperimenti dimostrò che il cap era costituito da **7-metilguanosina** (**m⁷G**) legata al nucleotide iniziale del trascritto mediante un insolito legame 5'-5' (**fig. 12.11**). La sequenza degli eventi che porta all'aggiunta del cap è illustrata in **figura 12.12** e richiede quattro diverse attività enzimatiche:

1. una RNA trifosfatasi rimuove il γ -fosfato all'estremità 5' del trascritto nascente;
2. una guanilil transferasi aggiunge un GMP, derivato da GTP, al difosfato rimasto all'estremità del trascritto formando un legame trifosfato 5'-5';

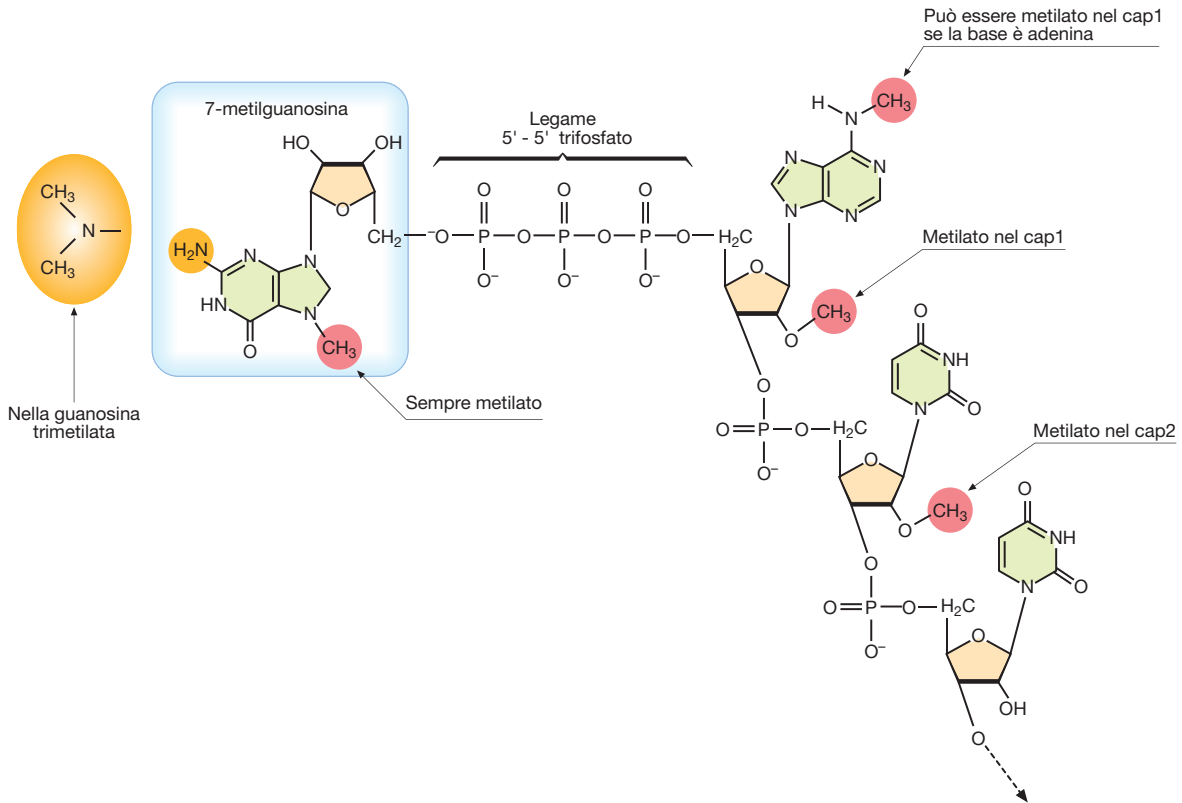
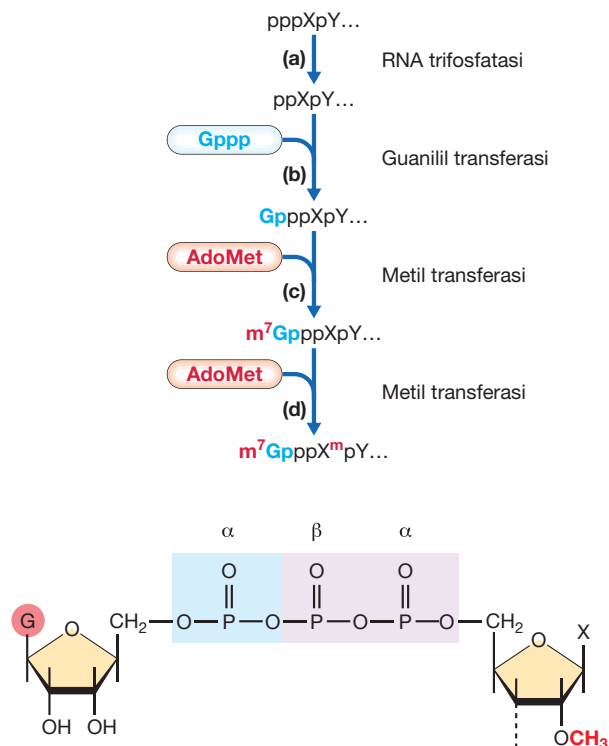


Figura 12.11 Struttura del cap presente all'estremità 5' degli mRNA eucariotici. Il cap consiste in una 7-metilguanosa legata con un insolito legame 5'-5' all'estremità dell'mRNA. Sono note

diverse forme di cap (cap1, cap2) in funzione della presenza di ulteriori siti di metilazione.

Figura 12.12 Sequenza degli eventi nel processo di capping. (a) L'RNA trifosfatasi rimuove il γ -fosfato dall'estremità 5' dell'mRNA; (b) la guanilil transferasi aggiunge il GMP, usando come substrato il GTP e liberando pirofosfato, per formare il legame trifosfato 5'-5'; (c) la metil transferasi aggiunge un gruppo metilico sulla guanina in posizione 7, utilizzando come substrato una molecola di S-adenosilmetionina (AdoMet); (d) ulteriori metilazioni possono dare luogo a forme varianti del cap (vedi fig. 12.11).



3. una RNA metil transferasi trasferisce un gruppo metilico dalla S-adenosil-metionina (AdoMet) all'azoto in posizione 7 della guanosina aggiunta nella reazione precedente;
4. un'altra metil transferasi usa un'altra molecola di AdoMet per metilare il 2'-OH del primo nucleotide del trascritto.

Si possono osservare diversi tipi di cap. Nel **cap1** risulta metilato il 2'-OH del primo nucleotide del trascritto, tipicamente una adenina, che a sua volta può essere metilata in posizione 6. Nel **cap2** risulta metilato al 2'-OH anche il secondo nucleotide del trascritto, mentre nel **cap0** non vi sono nucleotidi metilati oltre alla 7-metilguanossina. Il cap1 è presente sia in RNA virali che cellulari, il cap2 è presente solo in RNA cellulari, mentre il cap0 si osserva solo in alcuni RNA virali. Per alcuni RNA a basso peso molecolare (snRNA U1, snRNA U5 o snoRNA U3) si osserva una ipermetilazione del m⁷G cap, localizzata nel citoplasma, per formare 2,2,7-trimetilguanossina (**fig. 12.11**).

Il cap viene aggiunto nelle prime fasi della trascrizione, dopo che il trascritto nascente ha raggiunto la lunghezza di 20-40 basi, ma prima dell'occorrenza degli altri processi di maturazione quali splicing e poliadenilazione. Il posizionamento del complesso dei fattori proteici necessari per la formazione del cap al 5' del trascritto è assistito dal dominio C-terminale della RNA polimerasi II (CTD). Infatti, come abbiamo visto in altre sezioni del libro e come illustrato in **figura 12.13**, il CTD è coinvolto nei vari passaggi di maturazione del trascritto di Pol II: transizione da inizio ad allungamento della trascrizione, aggiunta del cap, splicing, taglio e poliadenilazione dell'estremità 3'.

In particolare, come schematizzato nella figura, nella transizione tra la fase di inizio e quella di allungamento della trascrizione, il CTD della RNA polimerasi II viene fosforilato da TFIIF sulla serina 5 dell'eptapeptide ripetuto YSPTSPS. Il CTD fosforilato (P-CTD) recluta il complesso proteico coinvolto nel capping e co-

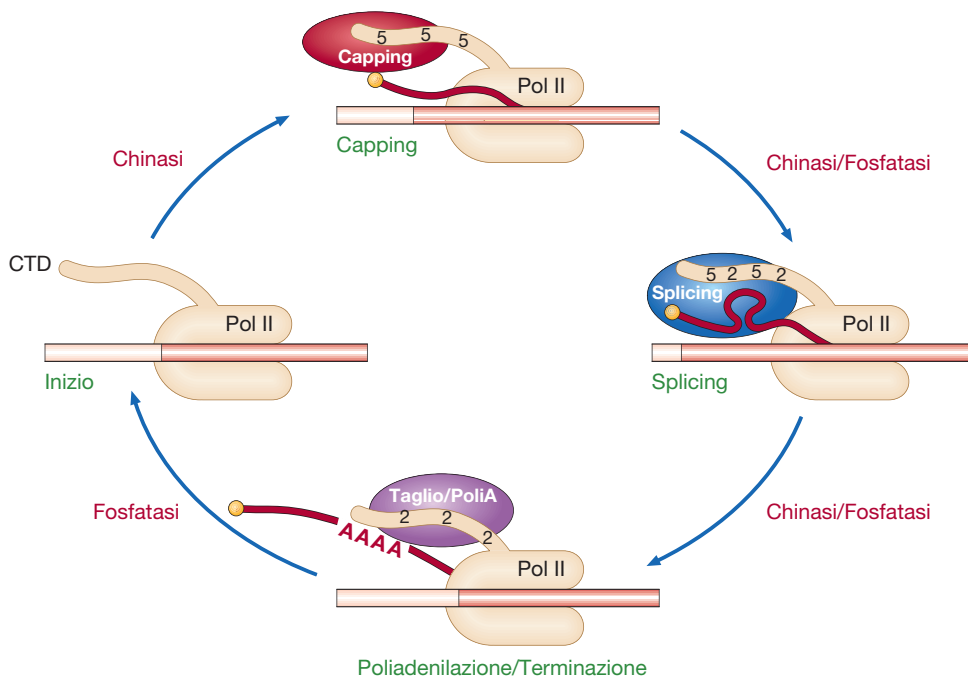


Figura 12.13 Ruolo della fosforilazione della coda CTD dell'RNA polimerasi II nel ciclo della trascrizione. Come osservato in varie parti del testo, il CTD della RNA pol II con i suoi stati fosforilati sulle serine 5 e 2 dell'eptapeptide ripetuto YSPTSPS, è coinvolto, oltre che nella transizione da inizio ad allungamento della trascrizione,

anche negli altri processi post-trascrizionali, e in particolare: l'aggiunta del cap, lo splicing, la terminazione e la poliadenilazione. I numeri 5 e 2 sul CTD della RNA pol II si riferiscono alla fosforilazione, rispettivamente della serina 5 e 2, dell'eptapeptide YSPTSPS.

stituito nei metazoi da una proteina bifunzionale (*capping enzyme*) con attività RNA trifosfatasi e guanilil transferasi e da una RNA metil transferasi. Questo spiega la specificità del capping rispetto ai trascritti sintetizzati dalla RNA polimerasi II.

Il cap svolge diverse funzioni:

- protegge l'mRNA dalla degradazione;
- aumenta l'efficienza della traduzione;
- facilita il trasporto dell'mRNA dal nucleo al citoplasma;
- contribuisce ad aumentare l'efficienza del processo di splicing.

Esperimenti condotti iniettando, in oociti di *Xenopus*, RNA di retrovirus dotati o meno di cap hanno dimostrato che i primi risultavano molto più stabili degli altri, indipendentemente dal fatto che la guanina al 5' fosse o meno metilata. Questo suggerisce che l'insolito legame 5'-5' tra il cap e il resto dell'mRNA sia resistente all'azione delle 5' esoribonucleasi, che invece possono degradare gli RNA privi di cap.

Un'altra importante funzione del cap è quella di rendere più efficiente la traduzione dell'mRNA. Infatti, il reclutamento del ribosoma sull'mRNA richiede il legame al cap di un fattore di inizio della traduzione denominato **eIF4E** (*cap-binding protein*), una proteina di 24 kDa che fa parte del complesso multiproteico eIF4F, responsabile del legame tra l'mRNA e la subunità minore del ribosoma (*vedi par. 15.2*).

La terza importante attività funzionale del cap è quella di facilitare il trasporto dell'mRNA maturo dal nucleo al citoplasma (*vedi par. 13.3*). Questa attività è stata dimostrata mediante esperimenti condotti su un piccolo RNA coinvolto nel processo di splicing, l'**snRNA U1**. Il gene che codifica per questo snRNA viene normalmente trascritto dall'RNA polimerasi II, producendo un trascritto dotato di cap. L'snRNA U1 così prodotto migra nel citoplasma, dove subisce il processo di trimetilazione del cap ($m^{2,2,7}G$), in seguito al quale ritorna nel nucleo, dove partecipa al meccanismo di splicing dell'mRNA precursore. Se si pone questo gene sotto il controllo dell'RNA polimerasi III, si ottiene un snRNA U1 privo di cap, incapace di traslocare nel citoplasma, provocando così anche il blocco del processo di splicing.

12.6 Poliadenilazione e terminazione della trascrizione degli mRNA eucariotici

Una caratteristica peculiare dell'mRNA eucariotico è la presenza di una coda di poli(A) all'estremità 3'. Quando l'RNA polimerasi II raggiunge la fine di un gene e trascrive uno specifico motivo di sequenza, denominato **segnale di poliadenilazione**, si innescano i meccanismi che portano alla terminazione del processo trascrizionale, al taglio del pre-mRNA e all'aggiunta di una coda di poli(A) alla sua estremità 3'. In particolare, la trascrizione del segnale di poliadenilazione attiva il reclutamento da parte della coda CTD dell'RNA polimerasi II di due complessi proteici denominati **Cleavage and Poly(A)denylation Specificity Factor (CPSF)** e **Cleavage stimulation Factor (CstF)**. Il segnale di poliadenilazione, che più propriamente può essere definito **segnale di cleavage o taglio**, è costituito normalmente dalla sequenza esanucleotidica **AAUAAA**, localizzata 5-30 nt a monte del sito di taglio e poliadenilazione. Sebbene la sequenza AAUAAA sia quella più comune, alcune sue varianti come AUUAAA o AGUAAA, anche se con minore efficienza, possono svolgere la stessa funzione.

Ci si poteva aspettare un meccanismo di poliadenilazione più semplice nel quale l'aggiunta della coda di poli(A) avvenisse all'estremità 3' del trascritto, una volta concluso il processo di trascrizione. In realtà questo non accade in quanto l'apparato di poliadenilazione non riconosce l'estremità 3' terminale del trascritto ma il segnale di poliadenilazione, mentre l'RNA polimerasi II è ancora nella fase di allungamento del trascritto. Naturalmente, la porzione extra del trascritto a valle del **sito di poliadenilazione**, corrispondente allo specifico nucleotide a cui viene aggiunta la coda di poli(A), viene successivamente degradata.

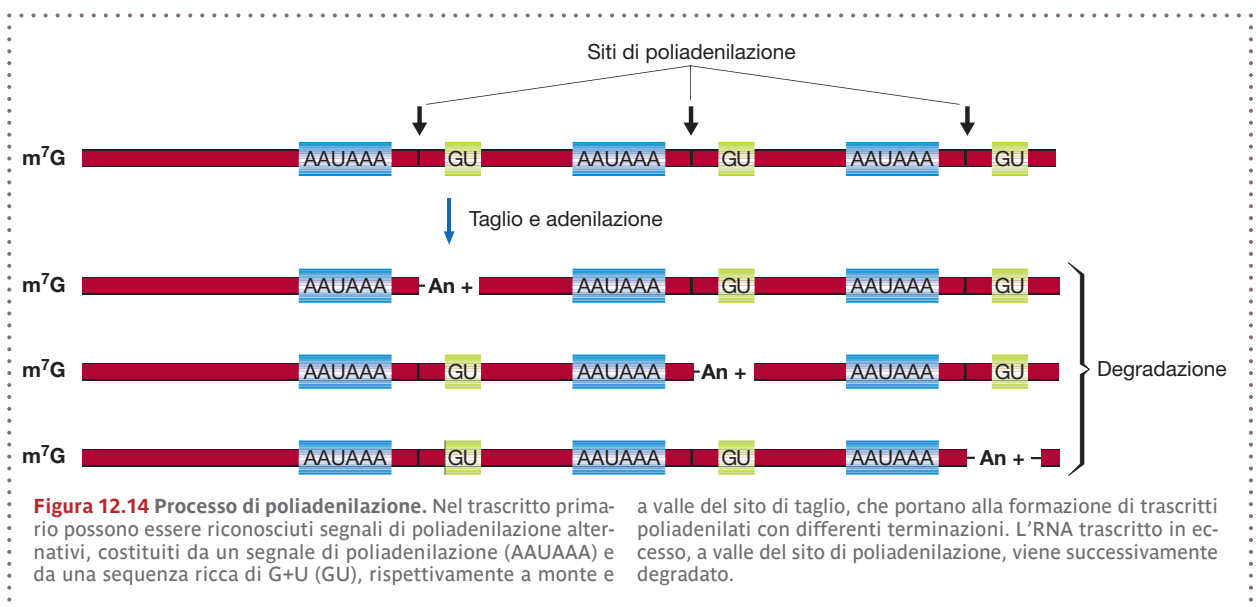
Tale meccanismo è differente da quello che si osserva nella trascrizione dei procarioti, in cui l'estremità 3' dell'RNA è determinata dalla terminazione della trascrizione.

L'apparato di poliadenilazione può riconoscere diversi segnali alternativi di poliadenilazione, formando, quindi, trascritti maturi poliadenilati di lunghezza variabile. Nel caso di siti alternativi di poliadenilazione, alcune evidenze sperimentali suggeriscono che la trascrizione si estende sempre oltre l'ultimo di questi siti liberando, in seguito a taglio e poliadenilazione, porzioni extra di RNA che verranno successivamente degradate (fig. 12.14).

È stato anche osservato che il segnale di poliadenilazione è necessario, ma non sufficiente, affinché il processo di poliadenilazione possa aver luogo in modo efficiente. Altrimenti, potremmo aspettarci molti siti spuri di poliadenilazione all'interno degli introni che sono generalmente ricchi in A+T. Infatti, 20-30 nt a valle del sito di poliadenilazione si osserva generalmente una sequenza ricca in G+U (o semplicemente ricca in U), che aumenta sensibilmente l'efficienza della poliadenilazione. Esperimenti *in vitro* con piccoli RNA sintetici hanno mostrato che la distanza ottimale tra il segnale di poliadenilazione e il sito di taglio dell'mRNA è pari a circa 8 nt.

Il segnale di poliadenilazione fin qui descritto è quello degli mRNA di mammiferi e vertebrati. Gli mRNA di piante e lieviti vengono anch'essi poliadenilati, ma posseggono segnali di poliadenilazione leggermente differenti. Mentre anche negli mRNA di piante il motivo AAUAAA è quello più rappresentato, anche se alcune sue varianti possono essere ugualmente efficienti, il segnale di poliadenilazione del lievito è molto meno conservato ed è normalmente costituito da una sequenza ricca in A+U.

L'apparato di poliadenilazione, come abbiamo già visto, comprende il complesso CPSF, che lega il segnale di poliadenilazione, e il complesso CstF, che invece lega la regione ricca in GU. Questi due complessi, ciascuno costituito da più subunità proteiche, agiscono in modo cooperativo insieme ad altri due fattori proteici, necessari per il taglio dell'mRNA, denominati **cleavage factor I e II (CFI e CFII)**, oltre naturalmente alla **poli(A) polimerasi** che catalizza l'aggiunta della coda di poli(A). Il fatto che solo i trascritti della RNA polimerasi II vengano poliadenilati, suggerisce che anche questa abbia un ruolo nello stimolare il processo di poliadenilazione. Effettivamente, è stato sperimentalmente dimostrato che questa attività vede il coinvolgimento della coda CTD dell'RNA polimerasi II.



La poliadenilazione avviene in due fasi, una fase di inizio – dipendente dal segnale di poliadenilazione – nella quale si ha l'aggiunta della prima decina di adenine e una fase di allungamento nella quale vengono rapidamente aggiunte ulteriori 200-250 adenine, che invece è indipendente dal segnale di poliadenilazione.

La fase di inizio della poliadenilazione richiede l'azione della **poli(A) polimerasi (PAP)** oltre che di CPSF. Nella fase di allungamento l'azione della PAP è invece coadiuvata da una *Poly(A) Binding Protein* denominata PABPN1 (*Poly(A) Binding Protein Nuclear 1*).

In **figura 12.15** è mostrato un ipotetico modello che illustra le diverse fasi del processo di taglio e poliadenilazione dell'mRNA, con l'indicazione dei fattori proteici coinvolti.

La struttura della poli(A) polimerasi è schematizzata in **figura 12.16**: essa contiene un dominio di legame per l'RNA (RBD), un dominio polimerasico (PM), due segnali di localizzazione nucleare (NLS1 e NLS2) e una regione ricca di serine e treonine (S/T).

Sono state isolate, tuttavia, varie isoforme di questo enzima, generate sia da splicing alternativo che da siti alternativi di poliadenilazione.

La **terminazione della trascrizione** della RNA polimerasi II risulta strettamente correlata al processo di taglio e poliadenilazione dell'mRNA appena illustrato. Essa avviene in due fasi. Nella prima, il trascritto in fase di sintesi, dopo aver legato il complesso proteico coinvolto nei processi di taglio e poliadenilazione, subisce una serie di tagli endonucleolitici a valle del sito di poliadenilazione mentre la trascrizione è ancora in corso nella regione di terminazione. Nella seconda fase, è proprio il taglio a livello del sito di poliadenilazione a dare alla RNA polimerasi il segnale specifico in seguito al quale essa si dissocia dallo stampo. Infatti, alcuni esperimenti condotti sia in lievito che in mammiferi hanno mostrato che il disaccoppiamento fra la terminazione della trascrizione e il taglio/poliadenilazione può essere causato da mutazioni che riguardavano fattori coinvolti nel processo di taglio a livello del sito di poliadenilazione, ma non quelli coinvolti nel processo di poliadenilazione.

Figura 12.15 Processo che porta alla formazione dell'estremità 3' dell'mRNA e alla sua poliadenilazione. Negli eucarioti la formazione dell'estremità 3' dell'mRNA viene generata da un taglio endonucleolitico a livello dell'mRNA in fase di allungamento. CPSF si lega al segnale di poliadenilazione AAUAAA, mentre CstF si lega alla regione ricca in GU (o in U) a valle del sito di poliadenilazione. A questo punto intervengono i fattori CFI e CFII che operano il taglio endonucleolitico. Infine, la poli(A) polimerasi (PAP) aggiunge la coda di poli(A) (in rosso) cui si lega una proteina specifica (PABPN1).

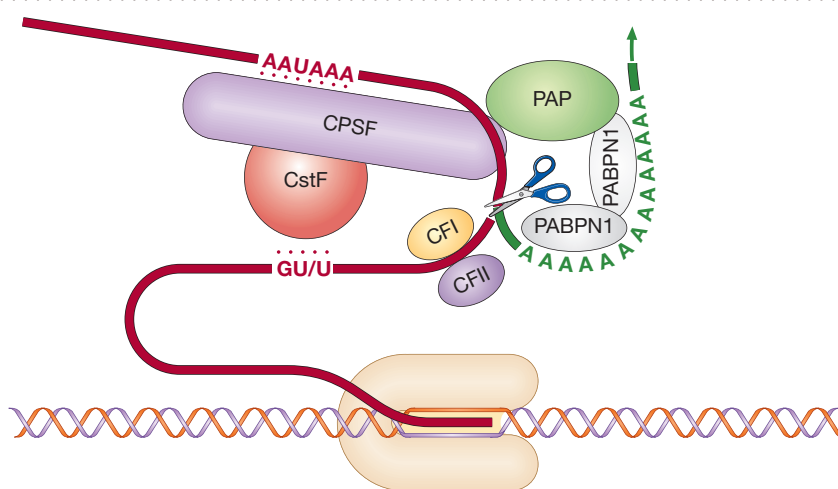


Figura 12.16 Struttura della poli(A) polimerasi (PAP). La PAP contiene un dominio di legame all'RNA (RBD), un modulo polimerasico (PM) responsabile dell'attività polimerasica, una regione C-terminale ricca in serina e treonina (S/T) e due domini di localizzazione nucleare (NLS1 e NLS2).



Funzioni della coda di poli(A)

La **coda di poli(A)** è presente nella maggior parte degli mRNA cellulari. Un'eccezione degna di rilievo è rappresentata dagli mRNA per le proteine istoniche, nei quali la coda di poli(A) è sostituita da una sequenza conservata in grado di assumere una struttura secondaria a stem-loop, localizzata nella regione 3' non tradotta dell'mRNA.

Come abbiamo già visto per il cap, la coda di poli(A) ha la funzione di stabilizzare l'mRNA, proteggendolo dalla degradazione, aumentare l'efficienza della traduzione, favorire i processi di splicing e della traslocazione nucleo-citoplasmatica.

In particolare, l'aumento dell'efficienza della traduzione dipende dal legame della PABPC1 (*Poly(A) Binding Protein Cytoplasmic 1*) alla coda di poli(A). È stato infatti mostrato che PABPC1, legandosi al fattore di inizio della traduzione eIF4G, induce la circolarizzazione dell'mRNA, che, favorendo il riciclo del ribosoma, rende molto più efficiente la traduzione dell'mRNA.

Circa l'azione della coda di poli(A) nel favorire il processo di splicing, evidenze sperimentali ottenute utilizzando substrati sintetici hanno mostrato che questa azione riguarda prevalentemente la rimozione dell'ultimo introne, ovvero quello più prossimo al sito di poliadenilazione.

La presenza della coda di poli(A) è necessaria anche per il trasporto attivo del trascritto maturo dal nucleo al citoplasma, dove fornirà lo stampo per il processo di traduzione.

Turnover della coda di poli(A)

La lunghezza della coda di poli(A) in un mRNA, una volta che questo sia stato traslocato nel citoplasma, non rimane costante ma può essere accorciata dall'azione di RNasi, contrastata dall'azione di una poli(A) polimerasi citoplasmatica che invece tende ad allungarla. Comunque, in genere si va verso l'accorciamento progressivo della coda di poli(A) a livello del citoplasma.

Il processo di poliadenilazione citoplasmatica è stato studiato in dettaglio durante il processo di maturazione degli oociti di *Xenopus*. L'oocita immaturo, infatti, contiene un gran numero di mRNA di origine materna, per lo più completamente deadenilati, e quindi traduzionalmente silenti. In seguito a uno stimolo a base di progesterone, alcuni mRNA vengono poliadenilati e poi efficientemente tradotti, mentre altri rimangono silenti e poi indirizzati alla degradazione.

Gli studi di Marvin Wickens e di Joel D. Richter hanno permesso di individuare un motivo a monte del segnale di poliadenilazione, costituito dalla sequenza **UUUUUAU** e denominato **CPE** (*Cytoplasmic Poly(A)denylation Element*), in grado di attivare il processo di poliadenilazione citoplasmatica. Descriveremo il meccanismo di questo controllo nel paragrafo 16.2.

Poliadenilazione nei procarioti e negli organelli

L'aggiunta della coda di poli(A) è un fenomeno presente non solo negli eucarioti, nei quali, come abbiamo appena visto, è un processo fondamentale per stabilizzare i trascritti e rendere più efficiente il processo di traduzione, ma anche nei procarioti e negli organelli, sia mitocondri che cloroplasti.

Tuttavia, a differenza di quanto osservato negli eucarioti, la poliadenilazione di procarioti e organelli è prevalentemente associata al processo di degradazione dell'RNA.

La degradazione dell'RNA poliadenilazione-dipendente normalmente richiede un taglio endonucleolitico, l'aggiunta di una coda di poli(A) (o ricca in A) e la successiva degradazione esonucleolitica dell'RNA. Un enzima chiave in questo processo è la **polinucleotide fosforilasi (PNPasi)**, che catalizza una fosforolisi processiva 3'→5' e una polimerizzazione 5'→3'.

Oltre alla PNPasi, responsabile della formazione di code eteropolimeriche ricche in A (%A > 70%), procarioti e organelli adottano anche delle poli(A) poli-

merasi (PAP), responsabili della formazione di code di poli(A) omopolimeriche.

L'abbondanza relativa di code omopolimeriche, generate dalla poli(A) polimerasi, ed eteropolimeriche, generate dalla PNPasi, varia in funzione dell'organismo. Per esempio, le code dei trascritti dei cloroplasti di spinacio risultano prevalentemente eteropolimeriche, mentre quelle di *Arabidopsis* risultano prevalentemente omopolimeriche.

Il pattern e la funzione della poliadenilazione degli RNA mitocondriali sono piuttosto eterogenei. Gli RNA mitocondriali di piante non appaiono costitutivamente poliadenilati, ed è stato suggerito che, come nei procarioti e nei cloroplasti, la poliadenilazione sia associata al processo di degradazione. Al contrario, i trascritti mitocondriali dell'uomo e degli altri animali sono stabilmente poliadenilati.

L'estrema compattezza del genoma mitocondriale dei metazoi fa sì che alcuni geni siano sovrapposti e talvolta contengano un codone di stop parziale costituito da U- o UA- che viene successivamente completato in Uaa o UAa proprio in seguito alla poliadenilazione.

Nei mitocondri umani sono stati identificati sia un omologo della PNPasi che uno della poli(A) polimerasi.

Il circuito di degradazione dei trascritti mitocondriali non è ancora stato elucidato. Tuttavia, come accade nei procarioti e negli altri organelli, anche nei mitocondri umani sono stati osservati trascritti tronchi e poliadenilati, tipici di un processo di degradazione dell'RNA, poli(A)-dipendente.

Alla luce di queste osservazioni si può supporre che la poliadenilazione accoppiata al processo di degradazione dell'mRNA sia una caratteristica primitiva, comune a tutte le forme viventi. Successivamente, i mitocondri dei metazoi e gli mRNA cellulari eucariotici, indipendentemente, hanno utilizzato questo processo con finalità differenti.

La scelta dell'adenina, e non di un altro nucleotide, può essere spiegata per la sua maggiore disponibilità, in quanto utilizzata anche dall'ATP, che costituisce la principale fonte energetica della cellula.

Lecture di approfondimento

- Blount, K.F., Uhlenbeck, O.C. (2005), «The structure-function dilemma of the hammerhead ribozyme», *Annual Review of Biophysics and Biomolecular Structure*, **34**, pp. 415-440.
- Cougot, N., van Dijk, E., Babajko, S., Séraphin, B. (2004), «Cap-tabolism», *Trends in Biochemical Sciences*, **29**(8), pp. 436-444.
- Cowling, V.H. (2009), «Regulation of mRNA cap methylation», *Biochemical Journal*, **425**(2), pp. 295-302.
- Dreyfus, M., Régnier, P. (2002), «The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria», *Cell*, **111**(5), pp. 611-613.
- Henras, A.K., Soudet, J., Gêrus, M., Lebaron, S., Caizergues-Ferrer, M., Mougin, A. e Henry, Y. (2008), «The post-transcriptional steps of eukaryotic ribosome biogenesis», *Cellular and Molecular Life Sciences*, **65**, pp. 2334-2359.
- Kim, V.N., Han, J., Siomi, M.C. (2009), «Biogenesis of small RNAs in animals», *Nature Reviews Molecular Cell Biology*, **10**(2), pp. 126-139.
- Mandel, C.R., Bai, Y., Tong, L. (2008), «Protein factors in pre-mRNA 3'-end processing», *Cellular and Molecular Life Sciences*, **65**(7-8), pp. 1099-1122.
- Mörl, M., Marchfelder, A. (2001), «The final cut. The importance of tRNA 3'-processing», *EMBO Reports*, **2**(1), pp. 17-20.
- Schuster, G. e Stern, D. (2009), «RNA polyadenylation and decay in mitochondria and chloroplasts», *Prog Nucleic Acid Res Mol Biol*, **85**, pp. 393-422.
- Srivastava, A.K., Schlessinger, D. (1990), «Mechanism and regulation of bacterial ribosomal RNA processing», *Annual Review of Microbiology*, **44**, pp. 105-129.

13

Splicing ed editing

13.1 Splicing

13.2 Editing

13.3 Traslocazione nucleo-citoplasmatica degli RNA

Questo capitolo è dedicato a due importanti processi di maturazione post-trascrizionale degli RNA, lo “splicing” e l’“editing”, che fanno sì che le sequenze nucleotidiche degli mRNA maturi siano diverse dalle sequenze dei precursori trascritti e quindi anche dalla sequenza dei geni corrispondenti. Ambedue i processi avvengono nel nucleo delle cellule eucariotiche, o negli organelli con proprio genoma quali mitocondri e cloroplasti, subito dopo la trascrizione. Sono invece assenti, tranne casi rarissimi, nei procarioti.

13.1 Splicing

Geni “discontinui” e splicing

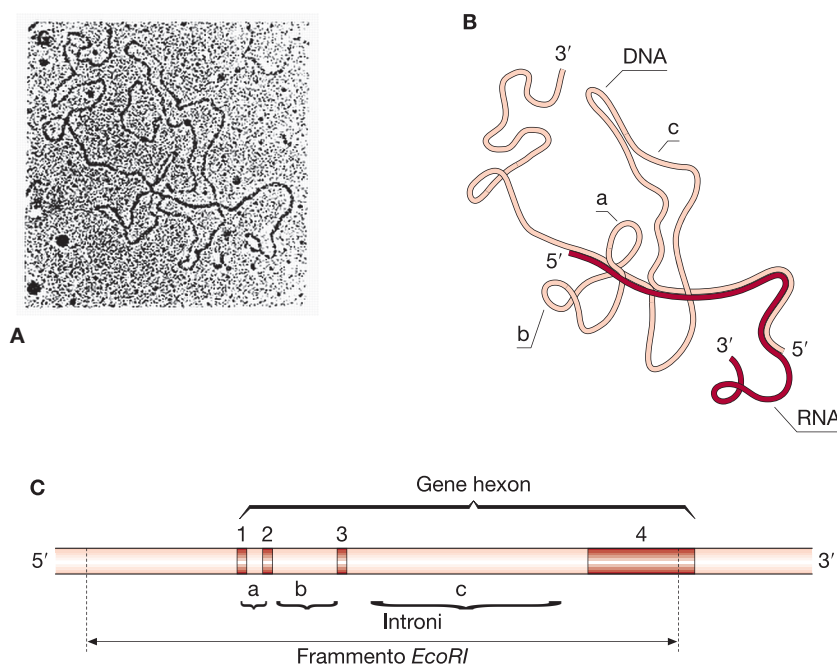
Alla fine degli anni settanta del secolo scorso, è stato dimostrato che la sequenza nucleotidica di un RNA funzionale non sempre è co-lineare con il DNA stampo. Infatti, a livello di DNA, la sequenza di un gene può contenere dei tratti di sequenza interposti che non compaiono poi nel trascritto maturo funzionale. Tali sequenze interposte sono state chiamate **introni**, mentre i tratti di sequenza che, unendosi, vanno a formare il trascritto maturo sono stati chiamati **esoni**. Geni di questo tipo sono stati chiamati **geni discontinui** (*split genes*). Abbiamo già parlato nel capitolo 5 (*vedi par. 5.3*) della organizzazione esoni-introni dei geni discontinui eucariotici e dei ruoli e del significato evolutivo degli introni.

La scoperta degli introni è stata fatta nel 1977 dai gruppi di Phillip A. Sharp e di Richard J. Roberts (**fig. 13.1**) mentre studiavano i geni di Adenovirus e poi estesa ad altri geni eucariotici cellulari da altri gruppi di ricerca. Le prime dimostrazioni della presenza degli introni sono state ottenute mediante una tecnica denominata **R-looping** nella quale si effettua una ibridazione tra l’RNA e il suo stampo di DNA. Questo esperimento è reso possibile dal fatto che, in presenza di un’elevata concentrazione di formammide, gli ibridi DNA-RNA hanno una temperatura di melting (*vedi par. 20.1*) superiore alla doppia elica DNA-DNA. Il risultato dell’ibridazione tra il DNA stampo a singolo filamento e l’RNA può essere poi visualizzato al microscopio elettronico. In **figura 13.2A** è mostrata un’immagine di microscopia elettronica relativa all’ibridazione tra un DNA a singolo filamento, contenente il gene *hexon* di Adenovirus, che codifica una proteina del capsido virale, con la regione a monte e l’mRNA maturo dello stesso gene. Sono chiaramente visibili tre anse (a, b e c in **figura 13.2B**) che corrispon-

Figura 13.1 Phillip A. Sharp e Richard J. Roberts. Il genetista e biologo statunitense Phillip Sharp (A) e il biochimico inglese Richard Roberts (B) sono stati i primi a dimostrare che i geni eucariotici sono discontinui per la presenza degli introni, scoperta per la quale hanno ricevuto il premio Nobel per la medicina nel 1993.



Figura 13.2 Dimostrazione della struttura esoni/introni del gene *hexon* di Adenovirus, ottenuta mediante analisi al microscopio elettronico di molecole ibride DNA-RNA. (A) Immagine al microscopio elettronico dell'ibrido tra l'mRNA *hexon* e un tratto di DNA a singolo filamento ottenuto da un frammento di restrizione *EcoRI-EcoRI* del DNA genomico di Adenovirus, che inizia a monte del gene *hexon* e finisce poco prima della sua terminazione. (B) Interpretazione dell'immagine in (A) in cui si vedono le anse di ssDNA relative ai tre introni (a, b, c) e i tratti di molecola ibrida DNA:RNA. (C) Struttura del gene *hexon* di Adenovirus.



dono a regioni di DNA non presenti nell'mRNA maturo e quindi agli introni. La struttura discontinua dei geni è stata poi confermata confrontando la struttura di un mRNA con quella del gene corrispondente, dapprima mediante mappe di restrizione e infine mediante sequenziamento.

Come già menzionato nel capitolo 5 (*vedi par. 5.3*), il numero di introni può variare tantissimo nei diversi geni e in differenti organismi. Nel lievito *Saccharomyces cerevisiae* solo pochissimi geni sono interrotti da introni, in genere uno, che normalmente hanno anche dimensioni piuttosto ridotte. Al contrario, nei geni umani il numero di introni può variare da nessuno (geni monoesonici o *intronless*) fino a oltre trecento (gene per la titina); in media i geni umani contengono 10-11 introni (e quindi 11-12 esoni) che sono mediamente più lunghi delle sequenze esoniche. Nei geni umani la lunghezza complessiva degli esoni rappresenta poco più del 5% di quella degli introni, che possono avere una lunghezza che può anche superare 1 megabase (ad es. il gene umano *ACCN1*). La **tabella 13.1** mostra le specifiche caratteristiche della struttura in esoni e introni dei geni di vari organismi eucariotici.

Gli introni non sono presenti solo nei geni per mRNA, ma talvolta anche in geni per tRNA e rRNA.

Tabella 13.1 Introni nei geni di alcuni organismi eucariotici. Le statistiche sono calcolate considerando un solo trascritto per gene, scegliendo quello di lunghezza maggiore. I dati sono tratti dalle relative banche dati di riferimento (Uomo, Topo, *Drosophila*, *C. elegans* da www.ncbi.nih.gov; Fugu da www.ensembl.org; *Arabidopsis* da www.arabidopsis.org; *S. cerevisiae* da www.yeastgenome.org).

Organismo	Numero di geni con almeno 1 introne	% geni con introni	Numero medio di esoni/gene	Lunghezza media mRNA (nt)	Lunghezza media totale degli introni	% regioni esoniche nel trascritto precursore
Uomo	17755	85	11	3273	60555	5
Topo	18311	79	10	2895	47718	6
Fugu	17876	97	12	1670	6058	22
<i>Drosophila</i>	11630	83	4	2173	4558	36
<i>C. elegans</i>	19558	95	6	1265	1705	43
<i>Arabidopsis</i>	22039	70	5	1580	911	71
<i>S. cerevisiae</i>	329	5	1	1352	340	99

Quando furono scoperti gli introni, si pensò che l'assenza, nella molecola trascritta matura, delle sequenze introniche di un gene poteva essere spiegata in due modi diversi:

1. il processo di trascrizione è discontinuo e l'RNA polimerasi quando giunge al termine di un esone "salta" a quello successivo;
2. il processo di trascrizione è continuo e produce un RNA precursore dal quale gli introni vengono rimossi successivamente in un processo di "taglia e cuci".

L'osservazione di una **classe eterogenea di RNA nucleari** ad alto peso molecolare, altamente instabili (*heterogeneous nuclear RNA*, **hnRNA**) avvalorò la seconda ipotesi. Infatti, gli hnRNA sono i precursori degli mRNA e nel nucleo subiscono il processo di splicing che comporta la rimozione degli introni. Inoltre, esperimenti di ibridazione tra il DNA codificante la β -globina di topo e sia l'RNA precursore che quello maturo hanno dimostrato che gli introni vengono prima trascritti e poi rimossi.

La **figura 13.3** illustra schematicamente la successione di eventi che portano da una sequenza genica (DNA) alla sequenza dell'mRNA maturo e poi, nel citoplasma, alla sequenza amminoacidica della proteina codificata.

In natura esistono vari meccanismi di splicing, che consistono di una serie ordinata di eventi di rottura e formazione di specifici legami fosfodiesterici al fine di ottenere una precisa estrusione degli introni. L'accuratezza del processo di splicing è particolarmente critica per le giunzioni esoniche che cadono all'interno della porzione codificante dell'mRNA. Infatti, lo spostamento di un sito di taglio anche di un solo nucleotide (e comunque di un qualunque numero di nucleotidi che non sia multiplo di tre) comporterebbe lo scivolamento della fase di lettura (*frame-shift*), con la conseguente sintesi di una proteina completamente differente da quella originaria.

Parleremo nei prossimi paragrafi di quattro diversi meccanismi di splicing, ma non è escluso che ne esistano altri non ancora identificati, varianti di questi o anche meccanismi completamente diversi. Il processo di splicing più diffuso è quello che si chiama **splicing nucleare**, specifico per i trascritti da Pol II, cioè per i precursori dell'mRNA. Questo processo richiede l'intervento di un grande complesso ribonucleoproteico, denominato **spliceosoma**. Parleremo poi di due meccanismi, chiamati **autosplicing**, in cui il processo può avvenire senza l'intervento di alcun fattore esterno essendo gli introni dotati di attività autocatalitica in grado di operare la loro autorimozione. Sia lo splicing nucleare che l'auto-splicing comportano due successivi eventi di trans-esterificazione per ottenere la

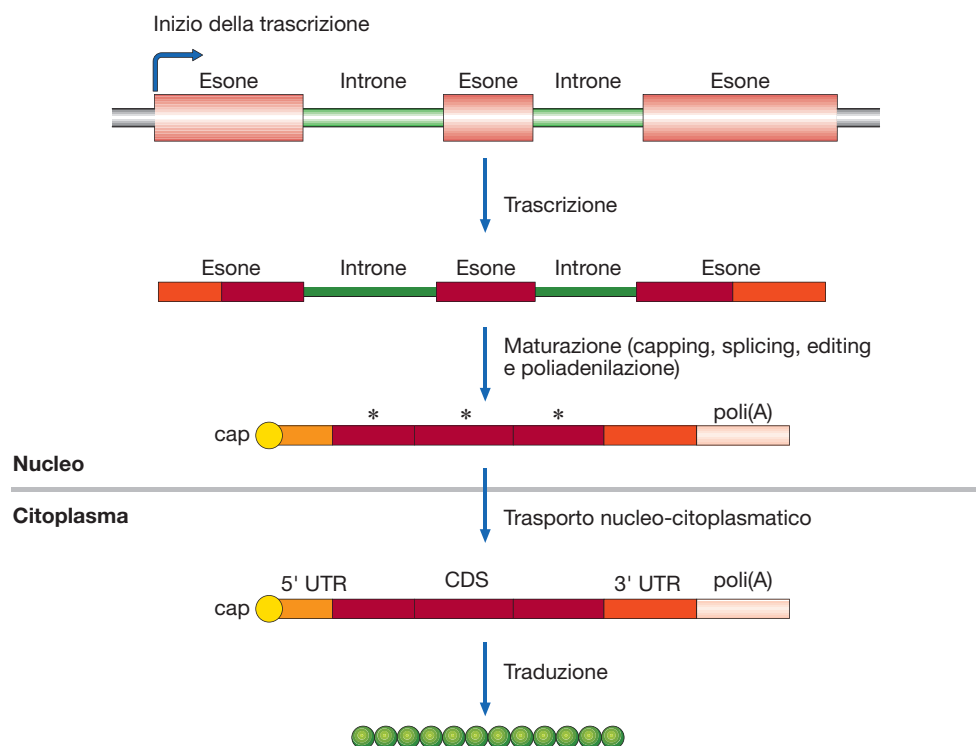


Figura 13.3 Processo di maturazione dell'mRNA. La maturazione dell'mRNA, che comprende capping, splicing, editing (indicato dagli asterischi) e poliadenilazione, precede la traslocazione dal

nucleo al citoplasma, dove avviene il processo della traduzione. Si noti la struttura ripartita dell'mRNA con le regioni non tradotte al 5' e al 3' (5'UTR e 3'UTR) e la regione codificante (CDS).

Tabella 13.2 I quattro meccanismi di splicing a confronto.

Donatore dell' – OH per la 1 ^a trans-esterificazione	Tipo di splicing	Catalizzatore
Adenosina interna all'introne (A)	Splicing nucleare	Spliceosoma
Guanosina esterna (exoG)	Splicing di introni di tipo II	Autosplicing
	Splicing di introni di tipo I	
Non ci sono trans-esterificazioni	Splicing dei tRNA di lievito	Nucleasi e ligasi

rimozione degli introni e la concatenazione degli esoni. Parleremo infine di un tipo di processo, non molto diffuso, in cui lo splicing viene ottenuto dall'azione combinata di nucleasi e ligasi, come esemplificato dal caso del processamento dei tRNA di lievito. La **tabella 13.2** mette a confronto alcune delle caratteristiche di questi quattro meccanismi di splicing.

Meccanismo dello splicing nucleare

Dopo la scoperta degli introni è stato subito osservato, per i geni trascritti da Pol II, che la sequenza nucleotidica degli introni, all'interno di uno stesso gene o in geni diversi, non è per nulla conservata e che il corretto riconoscimento dei siti di splicing nel precursore dell'mRNA è determinato da brevissimi segnali conservati in prossimità delle estremità delle sequenze introniche. In particolare è stato osservato che, come mostrato in **figura 13.4**, quasi tutti gli introni dei trascritti da Pol II iniziano con il dinucleotide GU e terminano con il dinucleotide AG.

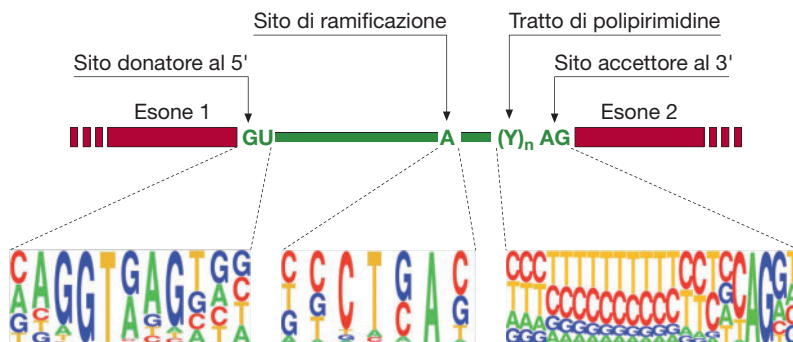


Figura 13.4 Tipica struttura di un introne da splicing nucleare. Il sito donatore GU al 5' e il sito di ramificazione contenente una A che si trova in prossimità di un tratto polipirimidinico (Y)_n che precede il sito accettore AG al 3'. Questi segnali, che permettono il riconoscimento e la rimozione degli introni, sono contenuti in motivi oligonucleotidici più estesi, rappresentati in basso, dove la dimensione del carattere corrispondente alla base è proporzionale al suo grado di conservazione (notare che qui il *sequence logo* è dato come DNA e quindi contiene T anziché U).

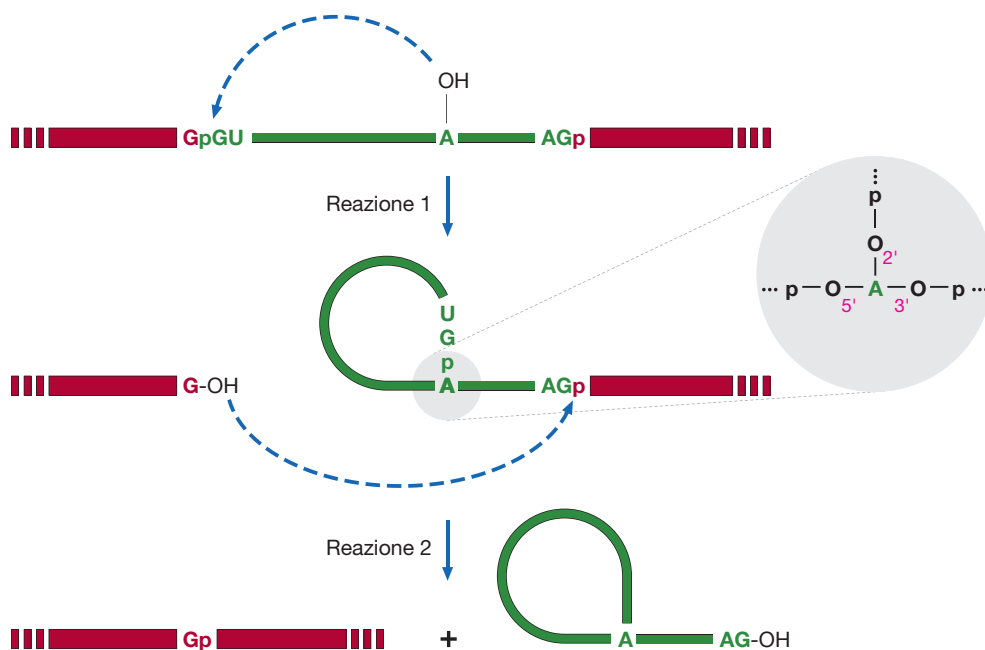


Figura 13.5 Le due reazioni di trans-esterificazione che hanno luogo nel processo di splicing nucleare. Il fosfato (p) tra i due esoni nel trascritto maturo (in rosso) è quello che si trovava al 3' dell'introne. La prima reazione di trans-esterificazione produce

un'adenina coinvolta in tre legami fosfodiesterici (ingrandimento mostrato a destra). La seconda reazione di trans-esterificazione concatena gli esoni e libera l'introne sotto forma di una struttura a forma di cappio (lariat).

Questi dinucleotidi conservati, a loro volta, sono all'interno di motivi oligonucleotidici più estesi che possono essere rappresentati da una sequenza consenso o da un *sequence logo* (vedi finestra 21.1). Oltre ai dinucleotidi conservati GU/AG che delimitano gli introni al 5' e al 3', che sono anche denominati, rispettivamente, **sito donatore** e **sito accettore**, negli introni vi sono altri due motivi conservati: una regione ricca in pirimidine subito a monte del sito accettore e un sito di ramificazione, anch'esso in prossimità dell'estremità 3' dell'introne (fig. 13.4).

La reazione di splicing nucleare comporta due successive reazioni di trans-esterificazione (fig. 13.5). La prima reazione consiste nell'attacco nucleofilo da parte del 2'OH dell'adenina presente nel sito di ramificazione al fosfato al 5' del dinucleotide GU. In questo modo si realizza un taglio al 5' della giunzione esone-introne che lascia un 3'OH libero all'estremità dell'esone; l'introne, che rimane legato all'esone a valle, assume una struttura a forma di cappio (lariat). Si può notare come l'adenina che ha dato inizio al processo risulta ora implicata in tre legami fosfodiesterici

che coinvolgono le posizioni 2', 3' e 5' del ribosio. Questa è la ragione per cui questa adenina viene denominata **sito di ramificazione** (*branching point*). La seconda reazione di trans-esterificazione consiste nell'attacco nucleofilo dell'OH libero all'estremità 3' dell'esone a monte sul fosfato al 5' dell'esone a valle. Questa seconda reazione completa il processo di splicing realizzando la concatenazione dei due esoni e il rilascio dell'introne sotto forma di struttura a forma di cappio, che verrà successivamente linearizzato da un enzima di deramificazione (*debranching*) e poi degradato. Entrambe le reazioni hanno una cinetica bimolecolare (S_N2), in quanto la loro velocità è correlata alla concentrazione di entrambi i substrati.

Il processo di splicing comporta, dunque, la rottura e la formazione di due legami fosfodiesterici. Pertanto la chimica di questo processo ha un bilancio energetico nullo, in quanto non richiede e non fornisce energia. Tuttavia, come si vedrà in seguito, il processo di splicing comporta un grande dispendio energetico, sotto forma di molecole di ATP, necessario per il corretto assemblaggio e funzionamento dell'apparato di splicing e per assicurare la specificità del riconoscimento delle sequenze introniche.

Un altro importante aspetto che va considerato riguarda la direzionalità della reazione. Considerando che il bilancio energetico della reazione è nullo, che cosa ci assicura che la reazione sia completamente spostata verso destra, ovvero nella direzione della completa rimozione degli introni e della concatenazione degli esoni? In realtà vi sono due elementi distinti che assicurano la direzionalità e l'irreversibilità della reazione:

- la reazione di splicing è favorita in quanto comporta un guadagno netto di entropia, dato che da una singola molecola di RNA precursore si ottengono più molecole, quali l'mRNA e gli introni che vengono rimossi;
- le molecole di RNA corrispondenti agli introni rimossi, e che costituiscono uno dei due prodotti della reazione, vengono rapidamente degradate e quindi non sono più disponibili per la reazione inversa.

■ Spliceosoma

Il processo di splicing nucleare richiede l'intervento di un grande complesso macromolecolare, costituito da oltre 200 proteine e 5 RNA, denominato **spliceosoma**. Tale complesso, la cui composizione può variare a seconda dell'organismo, ha dimensioni simili a quelle del ribosoma. Esso venne isolato in lievito nel 1985 da Edward Brody e John Abelson sotto forma di particelle con coefficiente di sedimentazione pari a 40S. Successivamente, lo stesso complesso, con coefficiente di sedimentazione pari a 60S, fu isolato da Phillip Sharp e collaboratori in cellule HeLa umane.

Lo spliceosoma rappresenta uno degli esempi più significativi di macchina molecolare in cui l'attività funzionale dipende dall'azione combinata di molecole di RNA e proteine. È interessante notare che i processi fondamentali dell'espressione genica come la maturazione dell'RNA ribosomale (*vedi* par. 12.1), lo splicing e la traduzione, implicano macchine molecolari di questo tipo.

Le 5 molecole di RNA presenti nello spliceosoma sono piccoli RNA nucleari ricchi in uridina (*small nuclear RNA*, snRNA) denominati U1, U2, U4, U5 e U6 (**tab. 13.3**), che nella cellula si trovano sotto forma di complessi ribonucleoproteici, denominati **snRNP** (*small nuclear RiboNucleoProtein*, pronuncia "snurp"), che contengono un snRNA, o due snRNA nel caso della snRNP U4/U6. La numerazione degli snRNA spliceosomali segue l'ordine temporale della loro scoperta; l'assenza di U3 è dovuta al fatto che è stato poi trovato che non è implicato nello splicing ma nella maturazione degli rRNA (*vedi* par. 12.1). Tutte le snRNP hanno un gruppo comune di sette proteine (**proteine Sm**) che si legano a un sito conservato negli snRNA, 5'-AAUUUGUGG-3', formando una struttura a ciambella, oltre a un gruppo di proteine specifiche per ciascuna snRNP.

Tabella 13.3 Elenco dei piccoli RNA nucleari (snRNA) dello spliceosoma. Per ciascun snRNA si riporta la lunghezza, l'identificativo della banca dati GenBank e la funzione.

snRNA	Lunghezza (nt)	GenBank ID	Funzione
U1	164	NR_004421	Riconosce il sito di splicing al 5' mediante l'appaiamento di una regione complementare
U2	187	NR_002716	Riconosce il sito di ramificazione mediante l'appaiamento con una regione complementare
U4	144	NR_003925	Forma un duplex con U6
U5	116	X04215	Funzione sconosciuta
U6	106	NR_004394	Forma un duplex con U4; scalza U1 nell'appaiamento con il sito di splicing al 5'

A differenza del ribosoma, la cui composizione rimane stabile nel corso della sintesi proteica, la struttura dello spliceosoma, come vedremo tra poco, è estremamente dinamica e ha una composizione variabile in funzione della specifica fase nel processo di splicing. Questa caratteristica rende possibile un livello molto elevato di regolazione che coinvolge numerose interazioni molecolari reversibili, quali RNA-RNA, RNA-proteina e proteina-proteina.

Lo spliceosoma ha sostanzialmente due funzioni principali:

- il corretto riconoscimento della coppia di siti di splicing tra una moltitudine di siti molto simili;
- il loro avvicinamento e corretto posizionamento nel sito catalitico per far avvenire in modo efficiente le due reazioni di trans-esterificazione.

Nei metazoi e nelle piante è presente anche uno **spliceosoma minore**, che ha in comune con quello maggiore l'U5 snRNA, ma che utilizza gli snRNA U11 e U12, U4atac e U6atac, funzionalmente analoghi a U1, U2, U4 e U6, rispettivamente, oltre naturalmente a uno specifico assortimento di proteine. Questo spliceosoma minore, anche denominato **U12**, è responsabile della rimozione di un numero piccolo di introni (poco più di un migliaio nell'uomo) ed è in grado di riconoscere sia i siti di splicing canonici GU/AG, sia siti non canonici come AU/AC e sue varianti. Gli introni rimossi dallo spliceosoma U12 contengono siti al 5' e al 3', così come il sito di ramificazione, molto più conservati di quelli dello spliceosoma maggiore, ma sono privi del tratto di polipirimidine tra il sito di ramificazione e il sito di splicing al 3' (**fig. 13.4**).

L'assemblaggio e il funzionamento dello spliceosoma seguono un ordine temporale progressivo ben preciso denominato **ciclo dello spliceosoma**.

Il ciclo inizia con il legame della snRNP U1 al sito di splicing al 5' attraverso un appaiamento complementare tra l'estremità 5' dell'snRNA U1 e il sito di splicing al 5' (**fig. 13.6A**). Quindi la proteina BBP (*Branch point Binding Protein*, anche detta SF1/BBP) si lega al sito di ramificazione e le due subunità del fattore U2AF si legano rispettivamente al tratto polipirimidinico (U2AF65) e al sito di splicing al 3' (U2AF35). Si forma in questo modo il **complesso E** (*Early*, precoce) che completa il riconoscimento dell'introne che deve essere rimosso.

A questo punto interviene la snRNP U2, che scalza la proteina BBP, utilizzando l'energia fornita da ATP, e forma mediante l'snRNA U2 un'interazione complementare con la regione del sito di ramificazione (*Branching Point*, BP) (**fig. 13.6B**). Questa interazione complementare non coinvolge la A del sito di ramificazione, che quindi protrudendo dalla regione di appaiamento viene resa competente per la prima reazione di trans-esterificazione. In questo modo si forma il **complesso A** o **pre-spliceosoma**.

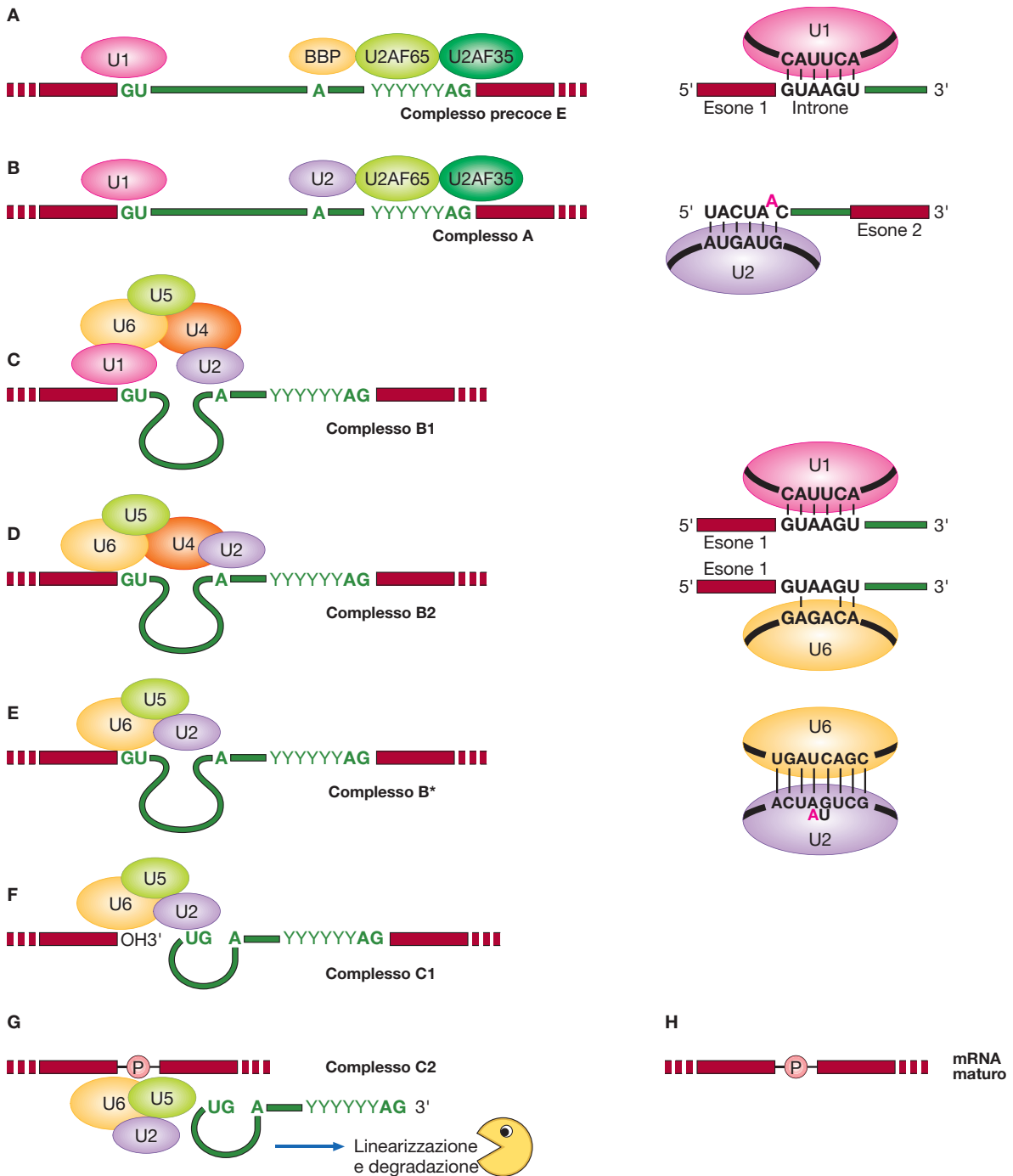


Figura 13.6 Meccanismo d'azione dello spliceosoma. (A) La snRNP U1 interagisce con il sito di splicing al 5' che presenta una regione di complementarità con l'snRNA U1 (vedi immagine a destra), mentre il sito di ramificazione e il sito di splicing al 3' interagiscono, rispettivamente, con le proteine BBP e U2AF. (B) La snRNP U2 scalza la proteina BBP sfruttando l'interazione complementare tra il sito di ramificazione e la snRNA U2 (vedi immagine a destra). (C) Il complesso ternario costituito dalle snRNP U4, U5 e U6, per mezzo di specifiche interazioni proteina-proteina, funge da ponte tra le snRNP U1 e U2 avvicinando i siti di splicing al 5'e

al 3'. (D) La snRNP U6 scalza la snRNP U1. (E) In seguito al rilascio della snRNP U4 si ha l'interazione tra le snRNP U6 e U2, per mezzo di regioni di complementarità tra i rispettivi snRNA. Si forma così lo spliceosoma attivato (complesso B*) che catalizza la prima reazione di trans-esterificazione che porta al complesso C1 (F). A questo punto si ha la seconda reazione di trans-esterificazione (G), che porta alla concatenazione degli esoni e al rilascio dell'introne sotto forma di struttura a forma di cappio (lariat) che viene successivamente linearizzato e degradato. L'mRNA maturo viene rilasciato e le snRNP riutilizzate per lo splicing di un altro introne (H).

Sul complesso A sopraggiunge la cosiddetta tri-snRNP, formata dalle snRNP U4/U6 e U5, che interagendo con le snRNP U1 e U2, induce l'avvicinamento tra il sito di splicing al 5' e il sito di ramificazione, favorendo così anche l'avvicinamento del sito di splicing al 3' immediatamente a valle. Si forma in questo modo il **complesso B1** (fig. 13.6C). Nel complesso B1 tutte le snRNP sono contemporaneamente presenti, ma esso è ancora cataliticamente inattivo.

Viene quindi rilasciata la snRNP U1, che viene rimpiazzata dalla snRNP U6 nell'interazione complementare con il sito di splicing al 5' (**complesso B2**; fig. 13.6D). Lo scambio tra U1 e U6 richiede la proteina Prp8, la più grande proteina spliceosomale, localizzata nella snRNP U5, insieme ad ATP. Successivamente viene rilasciata anche la snRNP U4, permettendo, in questo modo, l'interazione tra le snRNP U2 e U6 dovuta alla complementarità tra i rispettivi snRNA (fig. 13.6E).

Lo spliceosoma così attivato (**complesso B***) catalizza la prima reazione di trans-esterificazione, nella quale viene idrolizzata una molecola di ATP, formando il **complesso C1** (fig. 13.6F).

Sebbene il meccanismo catalitico non sia stato ancora del tutto chiarito, molte evidenze suggeriscono che gli snRNA appaiati U2 e U6 contengano tutte le principali componenti del sito attivo, utilizzando probabilmente come cofattore essenziale la proteina Prp8. È interessante notare che gli snRNA U2 e U6 presentano somiglianze strutturali con gli introni capaci di autosplicing del gruppo II. Pertanto, come vedremo in seguito per il ribosoma, è proprio la componente a RNA dello spliceosoma ad avere il ruolo principale nell'attività catalitica necessaria per il processo di splicing.

Nel complesso C1 l'introne forma un intermedio a forma di cappio (lariat) che viene poi staccato dall'esone a valle nella seconda reazione di trans-esterificazione, nella quale viene idrolizzata una seconda molecola di ATP (**complesso C2**; fig. 13.6G). Gli introni rimossi, sotto forma di lariat, vengono prima linearizzati e quindi degradati.

Infine, l'mRNA maturo viene rilasciato dallo spliceosoma che si dissocia nelle snRNP componenti, che vengono quindi riciclate in una successiva reazione di splicing (fig. 13.6H).

L'mRNA maturo non è nudo, ma associato a diverse proteine specifiche, denominate **mRNP** (*messenger ribonucleoprotein*), che svolgono un ruolo nel controllo dell'accuratezza dello splicing dell'mRNA, facilitano il trasporto dal nucleo al citoplasma e aumentano l'efficienza della traduzione. In particolare, alcune proteine che costituiscono il **complesso EJC** (*Exon Junction Complex*) si legano all'mRNA neoformato a circa 24 nt a monte delle giunzioni esone-esone, marcando così nell'mRNA maturo tutti i siti di splicing. La presenza delle EJC sull'mRNA spiega come la sua sorte nel citoplasma sia legata ai processi avvenuti precedentemente nel nucleo. Come vedremo nel paragrafo 16.3, la presenza delle EJC consente di riconoscere e indirizzare alla degradazione mRNA non funzionali generati da splicing incompleto o non corretto in base alla presenza di codoni di stop prematuri, localizzati almeno 50 nt a monte di un sito di splicing. Oltre ad avere un ruolo nel controllo di qualità degli mRNA maturi le EJC, attraverso l'interazione con altre proteine specifiche, possono anche modulare l'efficienza del processo traduzionale. È stato infatti osservato che mRNA espressi da un gene dotato di introni e che quindi hanno subito il processo di splicing, vengono tradotti in modo più efficiente di mRNA derivati da una copia dello stesso gene priva di introni.

Come per gli altri livelli dell'espressione genica, anche il meccanismo di splicing e la complessa macchina molecolare che lo attua non sono immuni da possibilità di malfunzionamento che sono alla base di numerose patologie, come illustrato nella **finestra 13.1**.

FINESTRA 13.1 SPLICING E PATOLOGIE

Il processo dello splicing dei pre-mRNA eucariotici ha il compito molto delicato e complesso di rimuovere in modo preciso, veloce ed efficiente le sequenze introniche. Come illustrato nel paragrafo 13.4, questo processo è mediato da una complessa macchina molecolare, lo spliceosoma, e da numerose proteine regolatorie che riconoscendo siti specifici sull'mRNA precursore favoriscono o reprimono specifici eventi di splicing. Infatti, malfunzionamenti di questo processo hanno conseguenze importanti che comportano l'insorgenza di diverse patologie, incluso il cancro.

Si può immaginare che, quanto maggiori sono il numero e la dimensione degli introni, o il numero delle possibili varianti di splicing, tanto più elevata sarà la probabilità di alterazioni nel processo a causa di mutazioni a livello dei siti coinvolti direttamente (siti donatore/accettore GU/AG e sito di ramificazione)

o indirettamente (enhancer e silencer dello splicing localizzati negli esoni e negli introni) nell'intero processo. Analogamente, il processo di splicing può essere difettivo a causa di mutazioni o per l'alterazione dell'espressione genica a carico di proteine dello spliceosoma o coinvolte nella regolazione dello splicing, come le proteine SR e hnRNP (par. 13.1).

In **figura 13F1.1** sono illustrati alcuni esempi di difetti dello splicing correlati a patologie. La caratterizzazione a livello molecolare delle alterazioni responsabili dell'insorgenza della patologia è la base per la messa a punto di strategie terapeutiche appropriate in grado di eliminare o attenuare gli effetti patologici. A questo scopo si possono utilizzare, per esempio, piccoli RNA antisense che, mascherando per complementarità le sequenze critiche per l'evento di splicing deleterio, possono ripristinare il processo di splicing normale e far regredire lo stato patologico.

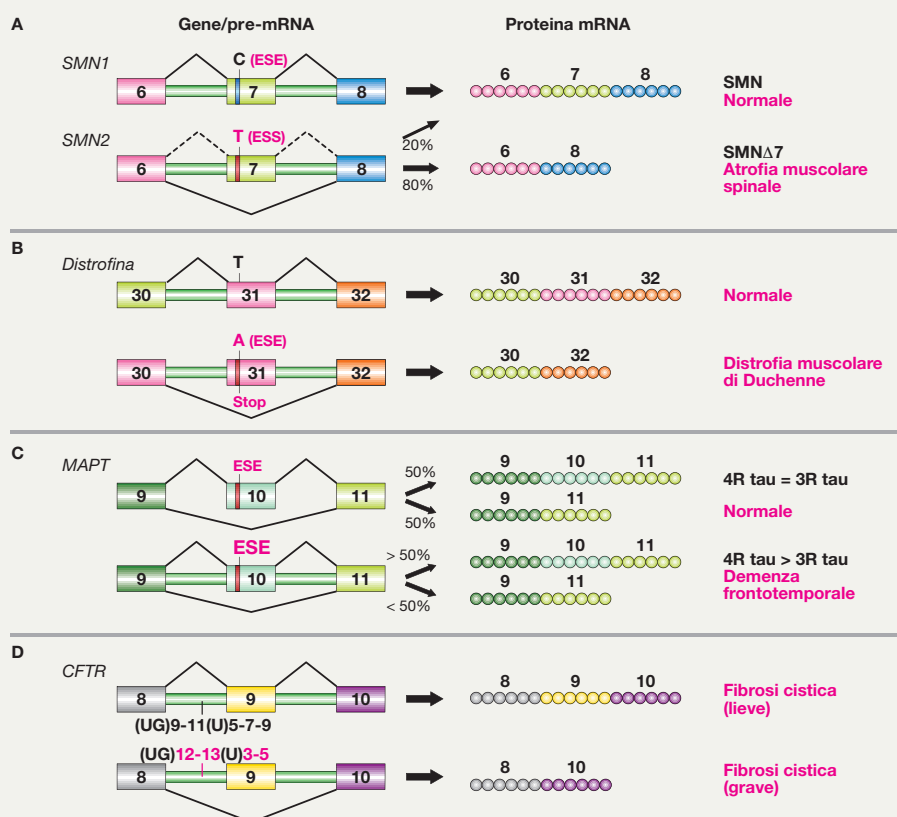


Figura 13F1.1 Esempi di difetti di splicing correlati a patologie. (A) L'atrofia muscolare spinale, detta anche SMA (*Spinal Muscular Atrophy*), è causata dalla delezione del gene *SMN1* che non viene compensata dall'espressione del gene *SMN2*, anche se questo è quasi identico al gene *SMN1*. L'eson 7 di *SMN1* differisce da quello di *SMN2* in un'unica posizione (la C in *SMN1* è sostituita da una T in *SMN2*). Questa sostituzione trasforma un enhancer esonico (ESE) in un silencer (ESS), con conseguente salto dell'eson 7, in circa l'80% dei pre-mRNA, e conseguente espressione di una proteina troncata non funzionale (*SMNΔ7*). (B) La sostituzione T/A nell'eson 31 del gene della distrofina genera allo stesso tempo un codone prematuro di stop e un ESS, con conseguente salto dell'eson 31. Questa mutazione induce una forma lieve di distrofia muscolare di Duchenne (DMD) in quanto l'mRNA privo dell'eson 31 produce una protei-

na solo parzialmente funzionale. (C) Mutazioni all'interno e a valle dell'eson 10 del gene *MAPT*, che codifica per la proteina tau, alterano alcuni elementi di regolazione dello splicing producendo uno sbilanciamento del normale rapporto 1:1 delle isoforme di mRNA con e senza l'eson 10. Questo comporta uno sbilanciamento delle proteine tau contenenti quattro o tre domini di legame ai microtubuli (4R-tau e 3R-tau, rispettivamente) che risulta nell'insorgenza della demenza frontotemporale. L'esempio riportato si riferisce alla mutazione N297K, che amplifica l'attività ESE, promuovendo l'inclusione dell'eson 10 e spostando l'equilibrio verso una maggiore concentrazione della forma 4R-tau. (D) Trattati polimorfici (UG)m(U)n presenti nell'introne 8 del gene *CFTR* influenzano il tasso di inclusione dell'eson 9 e il livello della proteina completa e funzionale, provocando fibrosi cistica con differenti gradi di gravità.

Autosplicing: introni di gruppo I e II

Nel 1982 Thomas R. Cech (**fig. 13.7**) e collaboratori scoprirono che un introne dell'rRNA del protozoo ciliato *Tetrahymena thermophila* era capace di autosplicing senza l'intervento di fattori proteici. Essi condussero un esperimento *in vitro* clonando, in un vettore plasmidico, una porzione del gene per l'rRNA della subunità grande del ribosoma contenente un introne. Ne effettuarono la trascrizione con una RNA polimerasi batterica purificata e inaspettatamente osservarono che il trascritto prodotto era in grado di effettuare autosplicing in presenza di ioni sodio e magnesio. Questa importante scoperta, per cui Thomas Cech ottenne nel 1989 il premio Nobel per la chimica, rivelò le proprietà catalitiche dell'RNA, che quindi poteva svolgere anche la funzione di un enzima (ribozima), e mise in luce un nuovo tipo di introni dotati di proprietà autocatalitiche.

Successivamente, questi introni sono stati raggruppati in due classi distinte e denominati **introni autocatalitici di classe I e II**, in funzione del meccanismo catalitico utilizzato nella reazione di splicing. Tuttavia, sia gli introni di classe I che quelli di classe II (anche detti di gruppo I e II) condividono la capacità di autosplicing, senza l'intervento di attività enzimatiche mediate da proteine, e non richiedono cofattori ad alta energia come i nucleotidi trifosfato.

Come avevamo già indicato in **tabella 13.2** e come descriveremo più in dettaglio poco oltre, anche i due meccanismi di autosplicing utilizzano due reazioni di trans-esterificazione, ma con una differenza sostanziale: mentre negli introni di tipo II la prima trans-esterificazione è effettuata, analogamente al processo di splicing nucleare catalizzato dallo spliceosoma, dal 2'OH di una adenina interna all'introne stesso, negli introni di tipo I essa è effettuata dal 3'OH di una guanosina esterna alla sequenza dell'RNA.

■ Introni di gruppo I

Gli introni del gruppo I, cui appartiene quello originariamente scoperto da Cech, hanno dimensioni generalmente comprese tra 250 e 500 nt e si possono trovare in una grande varietà di organismi eucariotici in geni nucleari, mitocondriali e plastidici, ma anche in batteri e batteriofagi.

L'analisi comparata di un gran numero di introni di gruppo I ha rivelato che questi assumono una struttura secondaria conservata che consiste in diverse regioni appaiate, numerate da P1 a P9 (o P10). L'analisi cristallografica di questa struttura ha mostrato che a livello tridimensionale essa è organizzata in modo da avvicinare i siti di splicing al 5' e al 3' ed è dotata delle attività catalitiche necessarie per le due reazioni di trans-esterificazione. In **figura 13.8A** è mostrata la struttura secondaria di un tipico introne di classe I, con l'indicazione delle eliche conservate P1-P9.

Il meccanismo di splicing, come abbiamo già detto, comporta due reazioni di trans-esterificazione, la prima delle quali effettuata da una guanosina esogena mono-, di- o trifosfato (exoG) il cui 3'OH effettua un attacco nucleofilo al fosfato al 5' dell'introne (**fig. 13.9A**). In questa reazione l'exoG viene legata all'estremità 5' dell'introne liberando il 3'OH dell'esone a monte. A questo punto ha luogo la seconda reazione di trans-esterificazione tra il 3'OH dell'esone a monte e il fosfato della guanosina all'estremità 3' dell'introne (ω G) che porta al rilascio dell'introne e alla concatenazione degli esoni.

■ Introni di gruppo II

Gli introni autocatalitici del gruppo II, con una lunghezza normalmente compresa tra 400 e 1000 nt, si trovano generalmente nei genomi mitocondriali e plastidici di protisti e piante, ma anche nei batteri. La maggior parte degli introni di gruppo II è costituita da due componenti principali: un ribozima in grado di catalizzare la reazione di autosplicing e una open reading frame codificante per proteine coinvolte nello splicing e mobilizzazione dell'introne (*vedi* oltre). Anche

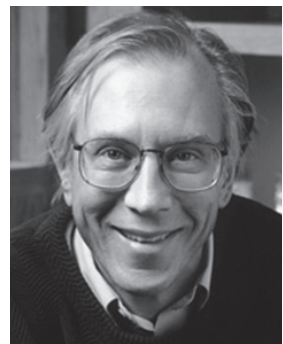


Figura 13.7 Thomas R. Cech. Il biochimico statunitense Cech (premio Nobel per la chimica nel 1989) ha scoperto, agli inizi degli anni ottanta, l'autosplicing dell'rRNA di *Tetrahymena*.

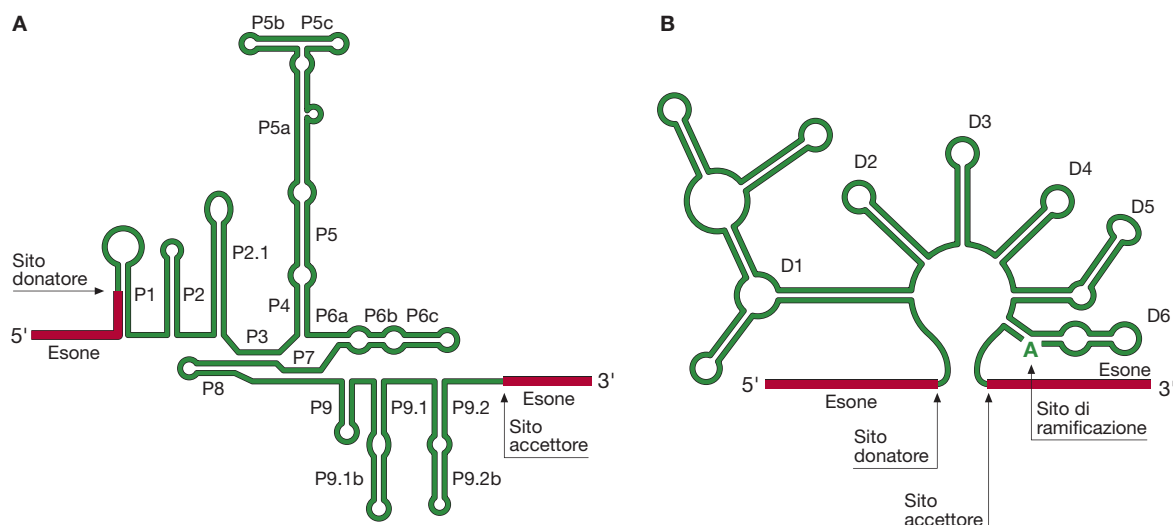


Figura 13.8 Tipica struttura secondaria degli introni da autosplicing. Le sequenze esoniche sono mostrate in rosso, quelle introniche in verde; le frecce indicano i siti di splicing al 5' (sito donatore) e al 3' (sito accettore). (A) Gli introni di gruppo I hanno una struttura secondaria conservata che consiste di nove regioni appaiate a forci-

na (P1-P9); queste poi si ripiegano a costituire essenzialmente due domini. (B) Gli introni di gruppo II hanno una struttura secondaria caratterizzata da sei tipiche strutture a forcina, chiamate domini 1-6, connesse da pochi nucleotidi nella regione centrale; il dominio 1 è il più grande ed è costituito da più sottodomini.

questi introni sono caratterizzati da una struttura secondaria conservata dotata di attività catalitica. La struttura secondaria del ribozima è caratterizzata da sei tipiche strutture a stem-loop, denominate D1-D6, che si irradiano da un nucleo centrale e si organizzano a livello tridimensionale in modo da portare in prossimità i siti di splicing al 5' e al 3' (**fig. 13.8B**). Si osservano pochissimi nucleotidi conservati, e in particolare a livello dei siti di splicing al 5' e al 3'.

Il meccanismo catalitico è analogo a quello degli introni spliceosomali e consiste di due reazioni di trans-esterificazione: la prima è data dall'attacco nucleofilo del 2'OH di un'adenina interna all'introne (nella regione D6) al fosfato del sito di splicing al 5', nella seconda il 3'OH dell'esone a monte effettua un attacco nucleofilo al fosfato nel sito di splicing al 3', risultante nella concatenazione degli esoni e nella formazione di un intermedio a forma di cappio (lariat) (**fig. 13.9B**).

La somiglianza tra il meccanismo di splicing degli introni di classe II e di quello operato dallo spliceosoma e l'osservazione di una significativa similarità strutturale della regione D5 con gli snRNA U6/U2 hanno portato a ipotizzare una correlazione evolutiva di questi due processi. L'attività spliceosomiale operante in *cis* in questa classe di introni autocatalitici si sarebbe evoluta per funzionare in *trans* negli introni spliceosomali.

Sebbene gli introni di gruppo I e II siano capaci di autosplicing *in vitro*, essi *in vivo* richiedono proteine accessorie che migliorino l'efficienza del processo di splicing, favorendo per esempio il loro ripiegamento nella struttura cataliticamente attiva. Alcune di queste proteine sono codificate da ORF localizzate negli stessi introni, generalmente in regioni non appaiate (dette anse o loop) della struttura secondaria, in modo da non interferire con la loro corretta conformazione strutturale. Le ORF localizzate negli introni di gruppo I hanno generalmente attività DNA endonucleasica, coinvolta nel meccanismo di retrotrasposizione (*Homing Endonuclease Gene*, HEG), ma possono anche avere un'attività di maturasi in quanto legandosi in modo specifico all'RNA precursore ne facilitano il processo di splicing. Gli introni di gruppo II contengono generalmente una ORF che codifica per una proteina con attività di trascrittasi inversa, coinvolta nella mobilizzazione degli introni, che può anche possedere ulteriori attività, come quella di maturasi precedentemente descritta per gli introni di classe I.

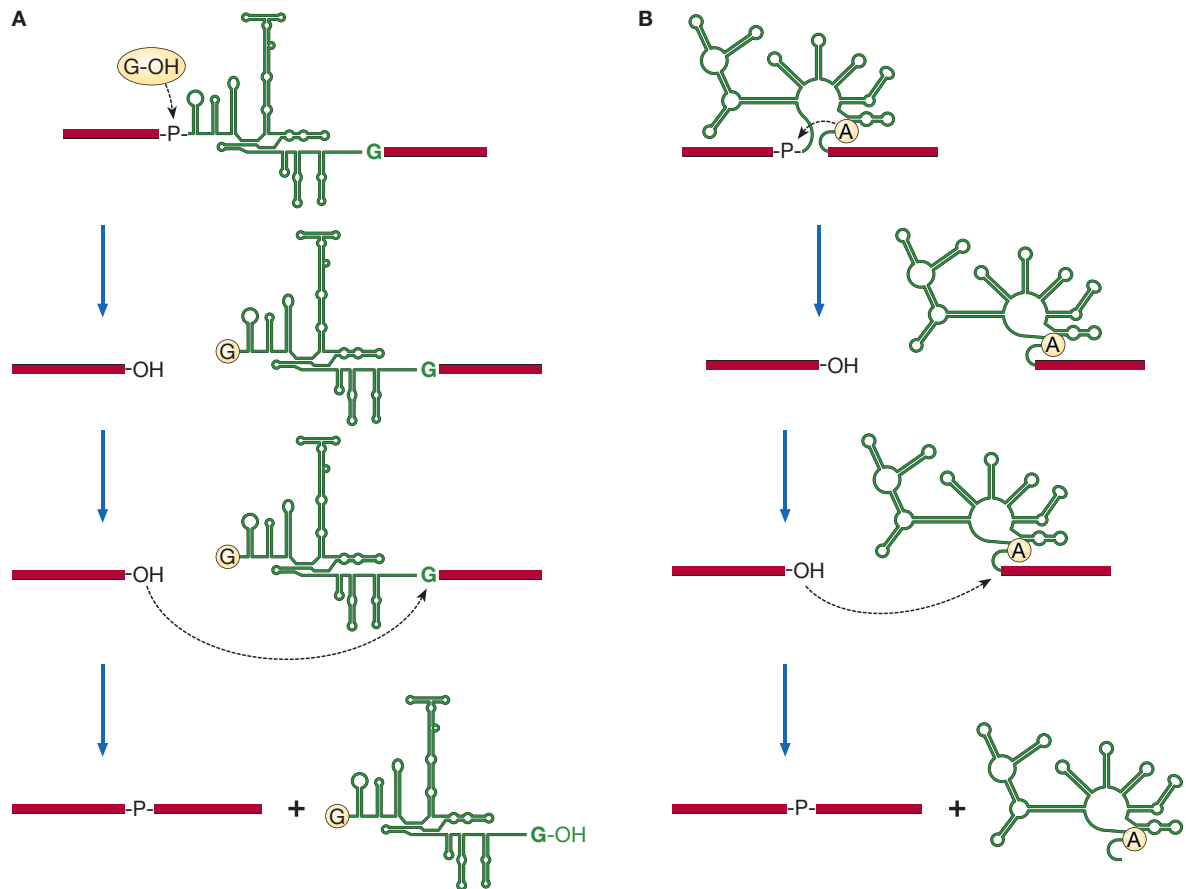


Figura 13.9 Meccanismi di splicing e di autocatalisi. (A) Autosplicing degli introni di gruppo I. L'attacco nucleofilo del 3'OH della guanosina (exoG nell'ovale in giallo) al fosfato al 5' del sito di splicing innesca il processo catalitico che induce la prima reazione di trans-esterificazione, la quale comporta il legame di una guanosina trifosfato libera (denominata exoG) al 5' dell'introne e taglia la giunzione al 5'. Il gruppo 3'OH dell'esone al 5' effettua quindi un

secondo attacco nucleofilo sul fosfato della guanine all'estremità 3' dell'introne (denominata ωG) che innesca la seconda reazione di trans-esterificazione, la quale porta al rilascio dell'introne e alla concatenazione degli esoni. (B) L'autocatalisi degli introni di gruppo II è analoga a quella degli introni spliceosomali; la prima reazione di trans-esterificazione è innescata dall'attacco del 2'OH dell'adenina nel punto di ramificazione.

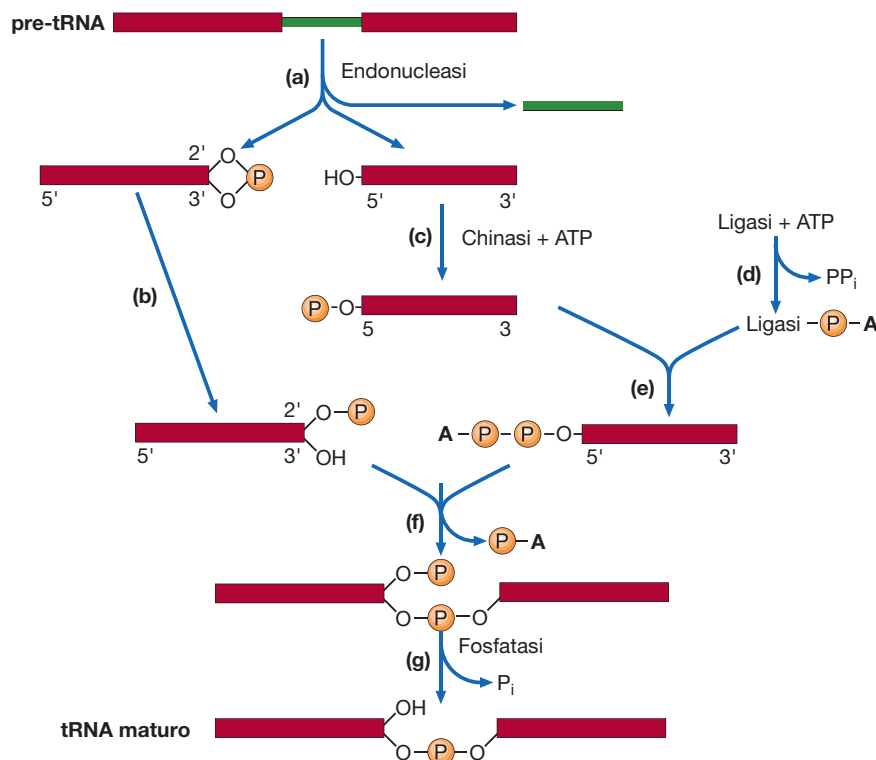
Gli introni del gruppo I e II possono anche essere considerati elementi genici mobili in quanto possono diffondersi in altre posizioni dello stesso o di altri genomi (trasferimento orizzontale) attraverso il processo di splicing e retrotrascrizione, un meccanismo anche denominato **splicing inverso**.

La scoperta dell'autosplicing, come anche dei processi di autotaglio e di varie altre attività catalitiche dell'RNA che abbiamo incontrato in varie parti del testo, ha suggerito che le forme di vita primordiali utilizzassero l'RNA sia come mediatore della trasmissione dell'informazione genetica che come strumento della sua espressione, essendo la sua attività funzionale mediata dalla sua attività catalitica. Tutte le forme di vita e i sistemi cellulari che oggi conosciamo avrebbero quindi avuto origine nel cosiddetto "mondo a RNA" (vedi par. 18.3).

Splicing del tRNA

Molti geni per tRNA in eucarioti e archeobatteri contengono introni che devono essere rimossi per generare i tRNA maturi funzionali. Il meccanismo di rimozione di questi introni, a differenza di quelli fin qui considerati, non comporta reazioni di trans-esterificazione ma l'intervento di diverse attività enzimatiche,

Figura 13.10 Meccanismo di splicing degli introni contenuti nei pre-tRNA. Una endonucleasi specifica rimuove l'introne effettuando due tagli alle sue estremità (a). Il tRNA risulta quindi costituito da due porzioni, di cui quella a monte termina con un fosfato ciclico che viene successivamente idrolizzato (b), mentre quella a valle viene fosforilata da una chinasi specifica (c). Interviene quindi una ligasi tRNA-specifica che completa il processo di splicing concatenando le due porzioni del tRNA (d-g).



tra cui endonucleasi e ligasi che operano in due fasi attraverso un meccanismo ATP-dipendente. La prima fase dello splicing è catalizzata da una endonucleasi specifica che riconosce una struttura conservata denominata BHB (*Bulge-Helix-Bulge*), effettua due tagli endonucleolitici che rimuovono l'introne dal pre-tRNA e produce il tRNA frammentato in due metà. La metà 5' termina con un fosfato ciclico 2'-3', lasciando un 5'OH all'estremità della metà 3'. Lo splicing viene quindi completato fondendo le due metà del tRNA attraverso l'azione di una ligasi tRNA-specifica (fig. 13.10).

Questo meccanismo di splicing, diffuso in un ampio spettro di specie negli archeobatteri e negli eucarioti, può essere utilizzato anche per la maturazione di mRNA cellulari, come nel caso del gene *XBPI*, coinvolto nei meccanismi di sorveglianza attivati dalla presenza di proteine con struttura disordinata (*Unfolded Protein Response*), nel quale viene effettuato uno splicing non convenzionale che rimuove un introne attraverso l'azione combinata di una endonucleasi sito-specifica e di una tRNA ligasi. Lo splicing non convenzionale di questo introne, generalmente di dimensioni molto ridotte (da 20 a 26 nt), è estremamente conservato nell'evoluzione in quanto si osserva negli omologhi del gene *XBPI* (HAC1 nel lievito) dal lievito all'uomo.

Trans-splicing

Le reazioni di splicing che abbiamo visto sin qui sono reazioni intramolecolari che producono la concatenazione di due esoni presenti nella stessa molecola di RNA precursore (cis-splicing). Tuttavia, lo splicing può anche aver luogo in *trans* quando le reazioni di trans-esterificazione sono intermolecolari e portano alla concatenazione di esoni provenienti da molecole differenti di RNA. Il tipo più comune di trans-splicing consiste nel trasferimento di un piccolo RNA non-codificante di lunghezza pari a circa 20 nt e denominato **Spliced Leader (SL)**, proveniente

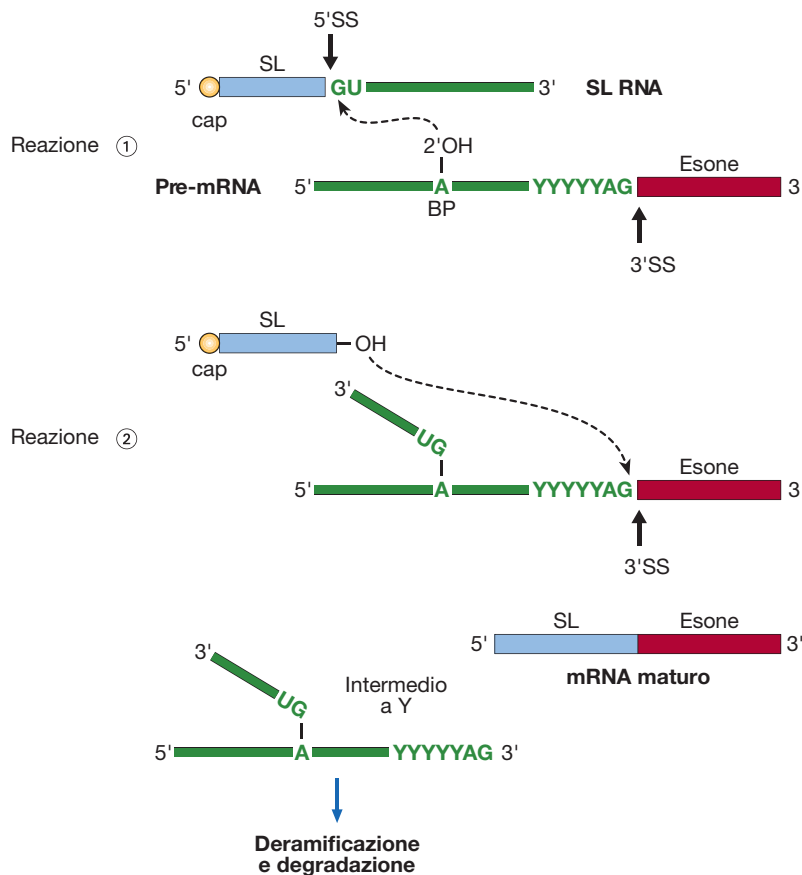


Figura 13.11 Il trans-splicing comporta due successive reazioni di trans-esterificazione tra due distinte molecole di RNA, un RNA SL e un pre-mRNA. L'SL RNA è dotato di cap (in arancio) e di un sito di splicing donatore (5'SS). Il pre-mRNA contiene una adenina al sito di ramificazione (BP, *Branching Point*), un tratto di poli-pirimidine (Y) e un sito accettore al 3' (3'SS). Nella prima reazione di trans-esterificazione il 2'OH dell'adenina nel BP effettua un attacco nucleofilo sul fosfato al 5'SS. Nella seconda trans-esterificazione l'attacco nucleofilo coinvolge il 3'OH dell'RNA SL e il fosfato al 3'SS. In questo modo si forma l'mRNA maturo dotato di un SL RNA al 5' con la liberazione di un RNA ramificato a forma di Y che viene rapidamente deramificato e degradato.

da un RNA donatore (SL RNA) contenuto in una snRNP, all'esone 5' terminale di un'altra molecola di RNA, trascritta indipendentemente. Come illustrato in **figura 13.11**, il trans-splicing adotta un meccanismo identico a quello già descritto per lo splicing nucleare e quindi richiede l'attività dello spliceosoma.

La differenza sostanziale sta nel fatto che il sito donatore (GU) e il sito accettore (AG) sono su due diverse molecole di RNA, con il sito di ramificazione e l'elemento ricco di pirimidine sull'RNA contenente il sito accettore.

L'aggiunta della sequenza leader, a seguito del trans-splicing, non contribuisce all'espansione della complessità del proteoma ma può svolgere una serie di importanti funzioni (*vedi oltre*).

Gli SL RNA sono trascritti dalla RNA polimerasi II, sono piuttosto corti (< 150 nt) e non presentano alcuna conservazione evolutiva a livello di sequenza o di struttura secondaria. Si trovano sotto forma di *Sm* snRNP e come gli snRNA U1, U2, U4 e U5 hanno un cap con trimetilguanosa (*vedi par. 12.5*).

Questo tipo di trans-splicing viene osservato in organismi unicellulari, ma anche in metazoi come i nematodi (*Caenorhabditis elegans*) e le ascidie (*Ciona intestinalis*), organismi nei quali i geni possono essere organizzati come negli operoni procariotici e vengono espressi come trascritti policistronici.

In *C. elegans* per il trans-splicing vengono utilizzati due distinti SL RNA, SL1 e SL2, che aggiungono entrambi un RNA leader di 22 nt. SL1, il più abbondante, viene aggiunto al 5' degli mRNA monocistronici e policistronici. SL2 viene invece aggiunto nelle posizioni interne dell'mRNA policistronico e contribuisce alla maturazione dell'mRNA policistronico risolvendolo in mRNA distinti monocistronici che, così dotati di cap, potranno produttivamente essere indirizzati al processo traduzionale (**fig. 13.12**). Infatti, l'aggiunta dell'RNA leader non solo

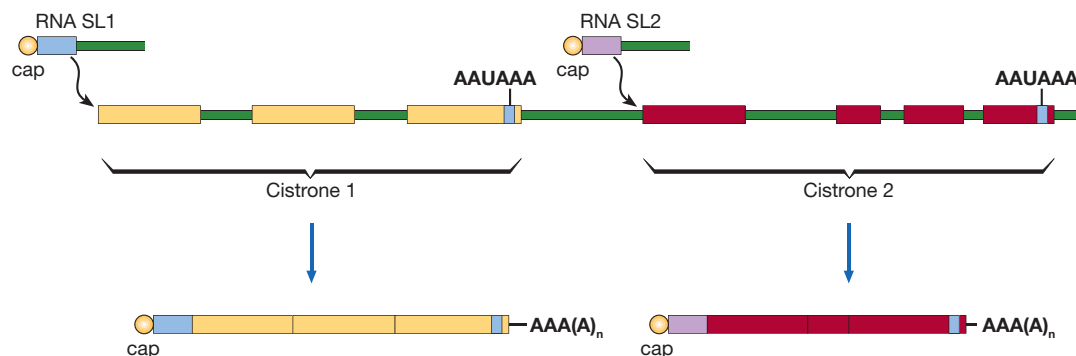


Figura 13.12 Risoluzione di un trascritto policistronico mediante trans-splicing in *C. elegans*. Il trans-splicing con l'RNA SL1 è specifico degli RNA espressi dal primo cistrone, mentre quello con l'RNA SL2 porta alla formazione di RNA monocistronici da parte

dei cistroni interni, che vengono così anche dotati di cap. Il processo di maturazione dell'RNA policistronico comprende anche il cis-splicing degli introni interni di ciascun cistrone e il taglio e la poliadenilazione a valle dei segnali di poliadenilazione AAUAAA.

è essenziale per la maturazione del pre-mRNA, ma attraverso l'aggiunta del cap aumenta in modo significativo l'efficienza della traduzione.

In *C. elegans* sono stati individuati circa 800 operoni costituiti ciascuno, in media, da 3-4 geni. Questi operoni contengono geni omologhi (derivati da eventi di duplicazione genica) o non omologhi, funzionalmente correlati e prevalentemente espressi in modo ubiquitario (ad es. proteine ribosomali, subunità del recettore dell'acetilcolina, enzimi che modificano il collagene, proteine coinvolte nello splicing).

Il trans-splicing si osserva anche per gli introni capaci di autosplicing. In questo caso, a differenza di quanto visto in precedenza, la porzione codificante del gene può essere suddivisa in due o più trascritti distinti dai quali può essere prodotto l'mRNA maturo in seguito a eventi di cis- e trans-splicing. La **figura 13.13** mostra schematicamente il processo di maturazione del trascritto primario codificante per la proteina nad2 nei mitocondri di *Arabidopsis thaliana*, che include sia cis- che trans-splicing oltre che RNA editing, un ulteriore meccanismo post-trascrizionale necessario a generare trascritti maturi funzionali che descriveremo più avanti.

Anche lo splicing del tRNA può avvenire in *trans*, come recentemente riportato nell'archeobatterio *Nanoarchaeum equitans* e illustrato in **figura 13.14**.

Vari esperimenti hanno dimostrato che il trans-splicing può riguardare anche gli introni spliceosomali dei geni nucleari di cellule animali. Sono stati osservati sia eventi di **trans-splicing intragenico**, attraverso il quale possono essere generate sequenze esoniche duplicate (ad es. in carnitina octanoil transferasi di ratto), che di **trans-splicing intergenico**, attraverso il quale possono essere generati mRNA chimerici distinti, concatenando esoni di trascritti distinti, anche codificati da geni localizzati su filamenti complementari [ad es. gene mod(mdg4) di *Drosophila*].

Approfondire e comprendere meglio il trans-splicing può avere importanti ricadute nella messa a punto di strategie innovative di terapia genica in cui questo meccanismo potrebbe essere utilizzato per correggere le alterazioni patologiche localizzate in una certa porzione del trascritto.

Accoppiamento tra splicing e trascrizione

Il processo di splicing è strettamente accoppiato al processo di trascrizione operato dalla RNA polimerasi II. Esso risulta, infatti, co-trascrizionale, dato che gli introni vengono rimossi dall'RNA in fase di allungamento via via che procede il processo trascrizionale. In questo modo è anche assicurato un più corretto riconoscimento del sito di splicing al 3', che potrà essere localizzato solo nella porzione di RNA già sintetizzata. Se consideriamo che un tipico gene umano contiene mediamente 8-10 introni, è evidente che se lo splicing fosse temporal-

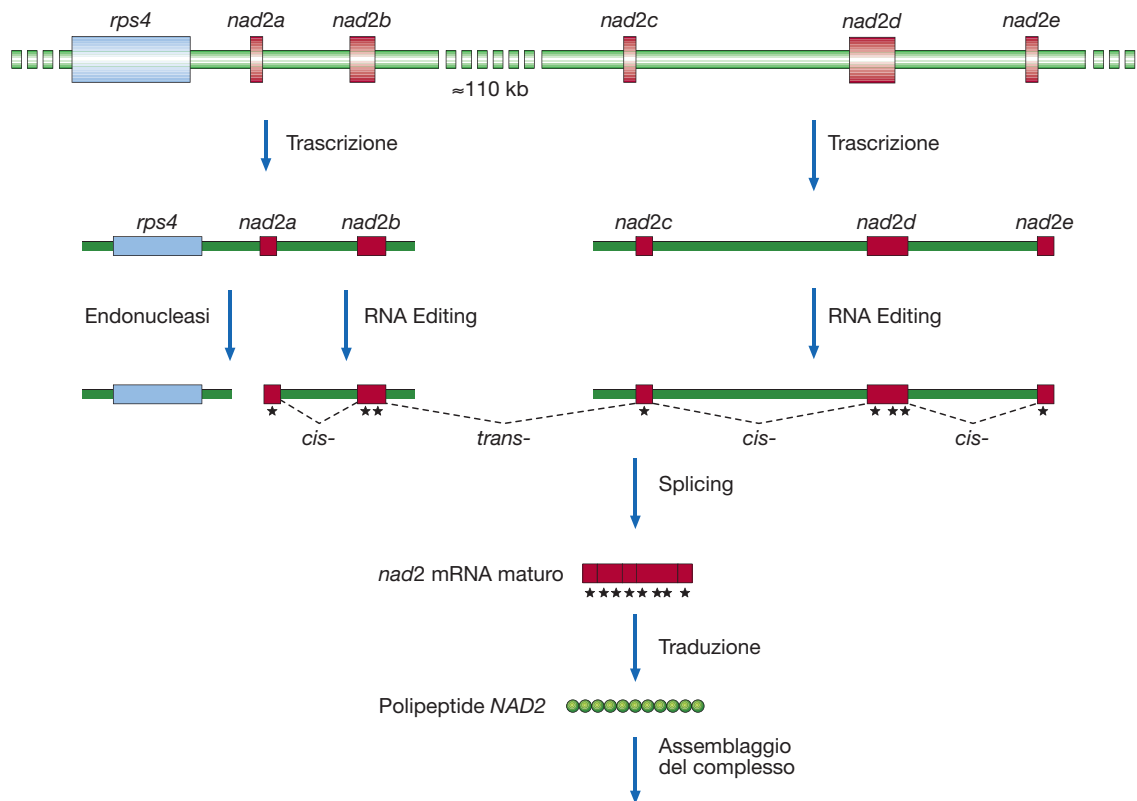


Figura 13.13 Maturazione del trascritto codificante per la proteina *nad2* nei mitocondri di *Arabidopsis thaliana*. L'RNA che costituirà il trascritto maturo è spezzettato in cinque esoni su due loci distinti del genoma mitocondriale (*nad2a-e*). Gli esoni a e b sono co-trascritti con un gene codificante per una proteina ribosomale (*rps4*). Un'attività endonucleasica è necessaria per ottenere il trascritto maturo di *rps4*. L'autosplicing in *cis* concatena gli esoni

ni dello stesso locus con la formazione di due frammenti di RNA (a+b, c+d+e) che vengono concatenati tra loro a opera del trans-splicing. L'RNA editing (vedi par. 13.2), rappresentato da asterischi, completa il processo di maturazione dell'mRNA modificando 31 basi. È importante precisare che i processi di splicing ed editing sono entrambi co-trascrizionali, e quindi avvengono simultaneamente via via che la trascrizione procede.

mente distinto dalla trascrizione vi sarebbe un'ambiguità molto maggiore nella selezione dei siti di splicing.

Anche in questo caso interviene la coda CTD dell'RNA polimerasi II. L'epta-peptide YSPTSPS, all'interno della coda CTD, già fosforilato in posizione 5 per reclutare i fattori proteici coinvolti nel capping, viene ulteriormente fosforilato in posizione 2 dal complesso della chinasi pTEFb per facilitare il reclutamento di fattori coinvolti nello splicing (vedi fig. 12.13).

Splicing alternativo

Una delle osservazioni più sorprendenti emerse dall'analisi di numerosi genomi eucariotici di cui è stata completata la sequenza nucleotidica è il fatto che non vi siano differenze significative nel numero di geni tra organismi con un livello di complessità molto diverso come il lievito (circa 6000 geni) e l'uomo (circa 20000-25000 geni) (vedi fig. 5.8). Questo apparente paradosso può essere almeno in parte spiegato dallo **splicing alternativo**, il meccanismo attraverso il quale uno stesso pre-mRNA può subire eventi di splicing differenti che portano alla creazione di diversi mRNA alternativi, che a loro volta possono codificare differenti proteine. Quindi, attraverso lo splicing alternativo, uno stesso gene può generare numerosi mRNA che presentano differenze sia nelle regioni 5' e 3' non tradotte (5' e 3'UTR) che nella regione codificante.

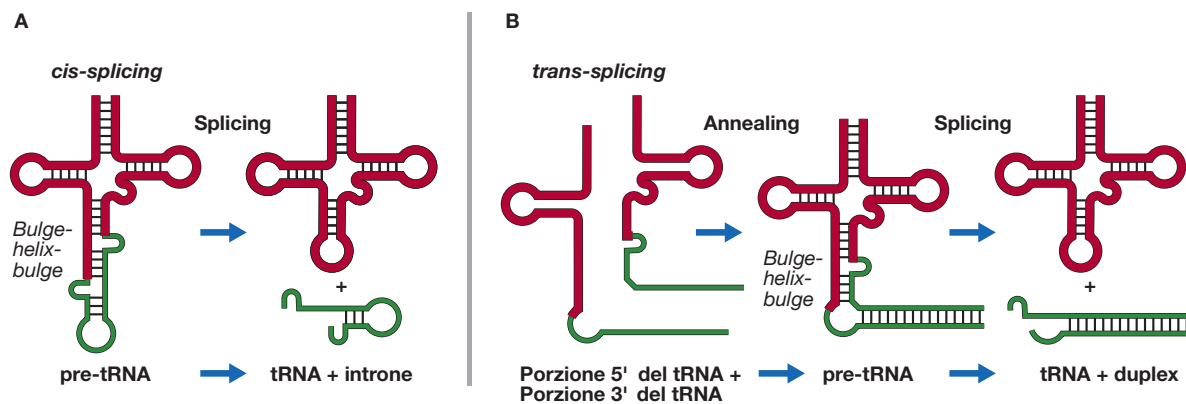


Figura 13.14 Rappresentazione schematica del cis- e trans-splicing negli introni dei tRNA dell'archeobatterio *N. equitans*. (A) Nel cis-splicing convenzionale del tRNA una specifica endonucleasi riconosce e taglia il motivo *bulge-helix-bulge* nel pre-tRNA, con conseguente escissione dell'introne (vedi anche fig. 12.4). (B) Nella fase iniziale del

processo di trans-splicing si ha l'appaiamento dei due frammenti che contengono le metà 5' e 3' del pre-tRNA. L'appaiamento coinvolge anche tratti delle regioni introniche presenti nei due frammenti del tRNA precursore (in rosso), facilitando la reazione di trans-splicing che avviene con lo stesso meccanismo del cis-splicing convenzionale.

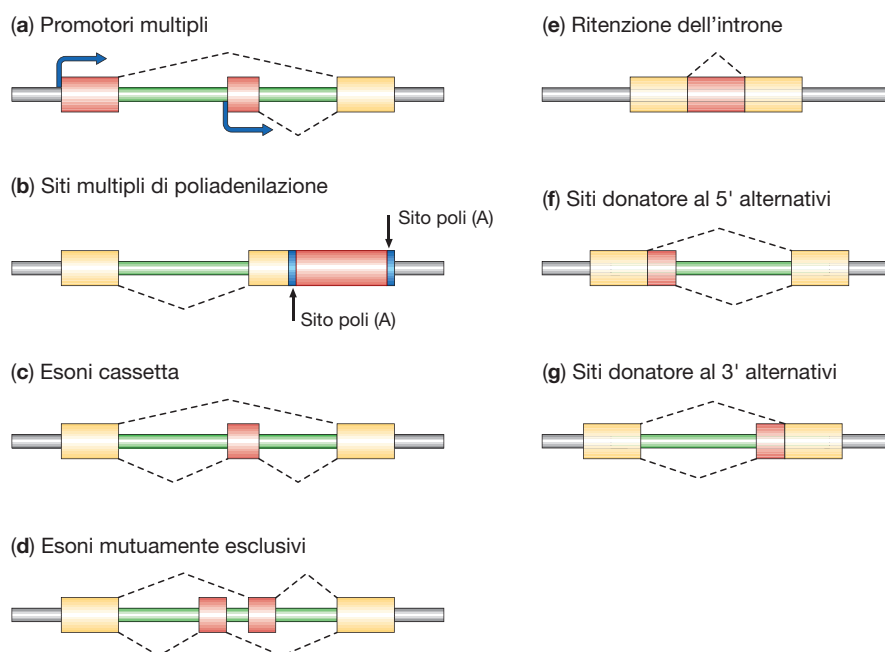
Studi recenti condotti utilizzando una piattaforma di ultrasequenziamento di seconda generazione (vedi par. 20.8) hanno dimostrato che oltre il 90% dei geni umani è interessato da splicing alternativo, che quindi rappresenta un processo chiave per espandere la complessità del trascrittoma e del proteoma.

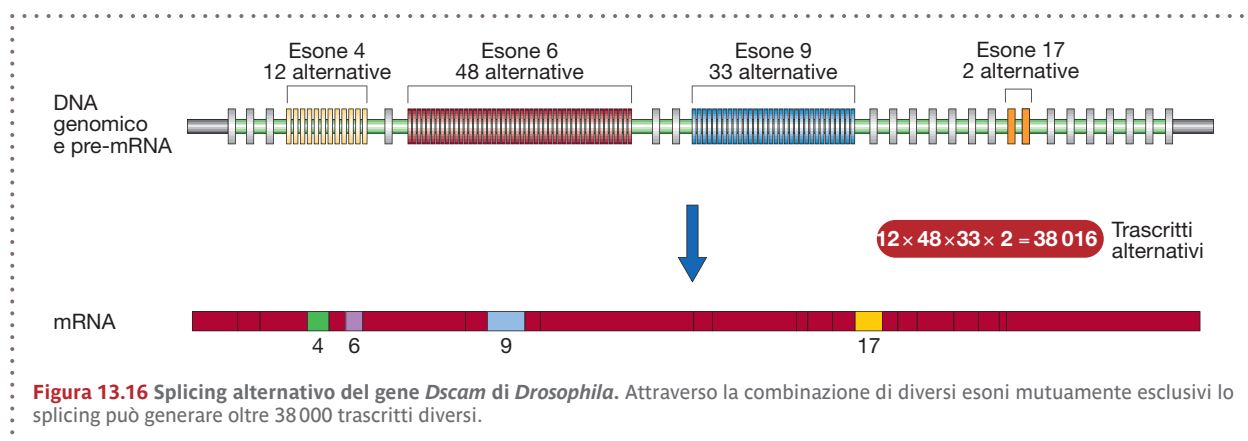
Lo splicing alternativo fu scoperto da David Baltimore e collaboratori, che osservarono che il gene per la catena pesante μ dell'immunoglobulina di topo poteva esprimere una forma secreta (μ_s) e una forma legata alla membrana (μ_m).

In **figura 13.15** sono illustrati tutti i possibili eventi elementari che concorrono alla generazione di più trascritti a partire da uno stesso gene. Questi includono:

- esoni facoltativi (detti anche esoni cassetta) ed esoni mutuamente esclusivi;
- ritenzione di introni;
- siti di splicing alternativi al 5' o al 3'.

Figura 13.15 Possibili eventi elementari che concorrono alla generazione di più trascritti a partire da uno stesso gene. Uno stesso gene può esprimere numerosi trascritti distinti attraverso l'utilizzo di siti alternativi di inizio (a) e terminazione (b) della trascrizione (AAUAAA è il tipico segnale di poliadenilazione), e lo splicing alternativo che comprende diversi eventi elementari (c-g).





Inoltre, l'uso di siti alternativi per l'inizio o la terminazione della trascrizione concorrono alla formazione di varianti trascrizionali di uno stesso gene.

Una variante trascrizionale di un determinato gene è, quindi, costituita da uno specifico assortimento di esoni (e introni), prodotta dalla combinazione di uno o più degli eventi elementari prima descritti e illustrati in **figura 13.15**.

L'assortimento combinatorio di eventi di splicing può portare alla generazione di un numero incredibilmente grande di trascritti, come nel caso del gene *Dscam* di *Drosophila*, codificante per proteine della superfamiglia delle immunoglobuline richieste per la formazione di connessioni neuronali, che può generare oltre 38 000 varianti, un numero maggiore del numero complessivo di geni di *Drosophila* (**fig. 13.16**).

Un esone presente in tutte le varianti trascrizionali è detto **costitutivo**, al contrario di un esone **facoltativo** (o **alternativo**) che viene o meno incorporato nella variante di splicing in funzione di specifici segnali di regolazione.

Regolazione dello splicing

Come abbiamo già visto, la maggior parte dei geni eucariotici è interrotta da introni. Lo splicing è quindi un processo fondamentale nell'espressione genica. Per quanto la presenza degli introni a prima vista possa apparire una inutile complicazione nel processo di espressione genica, essa in realtà introduce un incredibile livello di flessibilità alle potenzialità di espressione di un gene. La presenza degli introni, come abbiamo già visto nel capitolo 5, è molto vantaggiosa per due ragioni principali. Su scala evolutiva consente l'acquisizione da parte di una specie/popolazione di nuovi caratteri nel genoma attraverso il meccanismo dell'*exon shuffling* (vedi par. 5.3), che può generare nuovi geni o proteine formando nuove combinazioni di esoni. Inoltre, a livello individuale lo splicing alternativo consente un eccezionale livello di flessibilità nell'espressione genica che consente l'espressione di specifiche varianti con funzioni correlate, ma specializzate per diversi tipi cellulari, fasi dello sviluppo o del differenziamento, o per rispondere a specifici stimoli ambientali o condizioni patologiche.

Infatti, attraverso lo splicing alternativo si possono produrre trascritti differenti, utilizzando una diversa combinazione di esoni e introni, sia nella porzione codificante (dando quindi luogo a proteine diverse nella struttura e nella funzione) che nelle regioni non codificanti, 5'UTR e 3'UTR, così che le diverse varianti trascrizionali possano essere soggette a una diversa regolazione post-trascrizionale. È quindi evidente come lo splicing richieda un sistema di regolazione e controllo altamente specifico.

La regolazione dello splicing coinvolge elementi in *cis*, costituiti da specifici motivi sul pre-mRNA, alcuni dei quali, come i siti donatore, accettore e di rami-

ficazione, sono già stati descritti, ed elementi in *trans*, costituiti da fattori proteici che interagiscono in modo specifico con gli elementi di regolazione in *cis*. In pratica, tale meccanismo di regolazione rende possibile modulare l'efficienza del riconoscimento dei siti di splicing in modo che, a seconda della condizione e delle necessità cellulari, venga utilizzato uno specifico assortimento di siti di splicing così che vengano espresse le isoforme trascrizionali appropriate.

Sono stati ipotizzati due diversi meccanismi di riconoscimento dei siti di splicing a seconda delle caratteristiche degli introni e degli esoni presenti nei geni dei diversi organismi.

Il meccanismo basato sul riconoscimento degli esoni, che riconosce le coppie di siti di splicing a valle e a monte di un esone, è quello prevalente nel caso di geni, come quelli umani, con esoni relativamente piccoli e introni generalmente piuttosto grandi. Il meccanismo basato sul riconoscimento degli introni, al contrario, è quello prevalente nel caso di geni con introni relativamente piccoli quali quelli di invertebrati, piante e funghi. In questo caso vengono riconosciute le coppie di siti di splicing al 5' e al 3' di un introne.

In ogni caso, considerando anche che i motivi di sequenza caratteristici dei siti di splicing sono corti e degenerati, la possibilità di commettere errori, in seguito al riconoscimento di siti spuri, dovrebbe essere piuttosto elevata. Invece, si osserva che la fedeltà e la specificità dello splicing sono sorprendentemente molto alte. L'elevata specificità del riconoscimento dei siti di splicing dipende da motivi di sequenza addizionali, coinvolti nella regolazione dello splicing, e localizzati negli esoni o negli introni, che agiscono come attivatori (*enhancer*) o repressori (*silencer*) del processo di splicing. Pertanto, questi elementi di regolazione, in funzione della loro localizzazione all'interno del pre-mRNA e della loro attività, vengono classificati come **enhancer e silencer esonici** (ESE, *Exonic Splicing Enhancer*; ESS, *Exonic Splicing Silencer*) o **enhancer e silencer intronici** (ISE, *Intronic Splicing Enhancer*; ISS, *Intronic Splicing Silencer*). Questi elementi, che promuovono o inibiscono il riconoscimento dei siti di splicing sotto il loro controllo, sono stati identificati attraverso esperimenti di mutagenesi sito-specifica, analizzando l'effetto di specifiche mutazioni prodotte sperimentalmente, o mediante analisi bioinformatiche.

Gli elementi ESE vengono normalmente riconosciuti da proteine della famiglia SR che contengono uno o più domini in grado di legare l'RNA e una regione, detta SR, ricca di serine (S) e arginine (R). L'interazione tra ESE e proteine SR facilita l'assemblaggio dello spliceosoma.

Gli elementi ESS vengono, invece, legati da repressori dello splicing, costituiti da una famiglia di ribonucleoproteine denominate hnRNP (*heterogeneous nuclear RiboNucleoProtein*) contenenti uno o più domini RRM (*RNA Recognition Motif*) in grado di legare l'RNA. L'interazione tra i motivi ESS e le proteine hnRNP può inibire lo splicing attraverso vari meccanismi, che tipicamente consistono nell'inibizione di interazioni essenziali tra le componenti dello spliceosoma (ad es. hnRNPI impedisce l'interazione tra le snRNP U1 e U2) o nel mascheramento degli stessi siti di splicing.

Gli elementi di regolazione intronici, ISE e ISS, hanno un meccanismo di azione simile, ma sono stati meno studiati degli elementi esonici.

Gli elementi di regolazione dello splicing normalmente agiscono in modo cooperativo e sono lo specifico assortimento e concentrazione di repressori e attivatori (**fig. 13.17**) e la loro localizzazione a determinare i pattern di splicing specifici che si osservano in diversi tipi cellulari, stadi di sviluppo o condizioni fisiologiche. In qualche caso si è anche osservato come uno stesso elemento di regolazione possa agire sia da enhancer che da silencer in funzione della sua specifica localizzazione nell'esone o nell'introne (**fig. 13.18**). A parte il contesto in cui sono localizzati gli elementi di regolazione dello splicing, che ne determinano l'attività specifica, un altro aspetto che va considerato è la struttura secondaria del pre-mRNA, che può influenzare l'attività dei vari elementi di regolazione e anche

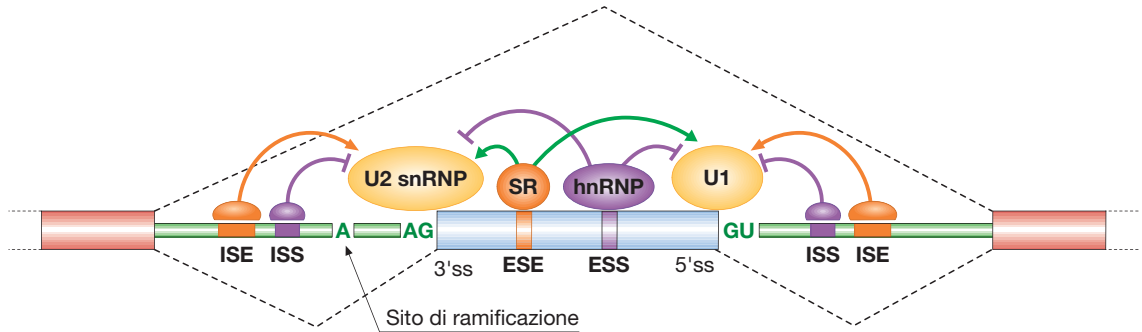


Figura 13.17 Regolazione dello splicing. La regolazione dello splicing comporta l'interazione tra proteine in grado di legare specifici motivi presenti nel pre-mRNA sia negli esoni (ESE, ESS) che negli introni (ISE, ISS). All'azione delle proteine SR che legano le sequenze enhancer negli esoni (ESE) e negli introni (ISE), e che at-

tivano lo splicing (freccie arancioni e verdi a punta), si oppongono le proteine hnRNP, che invece riconoscono le sequenze silencer negli esoni (ESS) e negli introni (ISS) (freccie viola a punta piatta). Il meccanismo di attivazione o repressione normalmente comporta l'interazione con altre componenti dello spliceosoma.

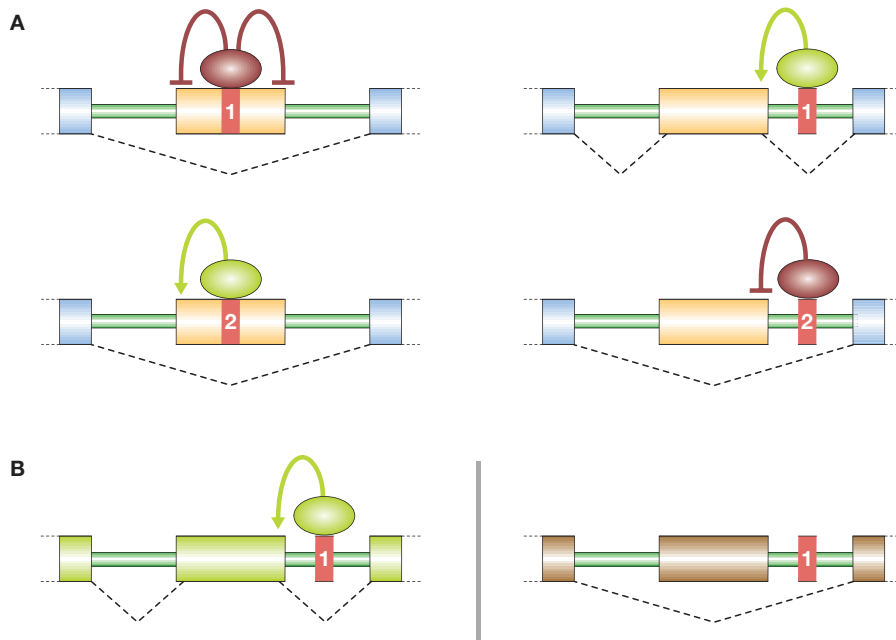


Figura 13.18 L'attività dei motivi di regolazione dello splicing è dipendente dal contesto. (A) Lo stesso motivo (1, 2) può fungere sia da attivatore che da repressore in funzione della sua localizzazione esonica o intronica. (B) Lo stesso motivo (1), anche nella stessa localizzazione intronica, può avere attività differenti in diversi geni (mostrati con un colore diverso). Nel gene mostrato nel pannello a destra, il motivo di regolazione, che normalmente agisce come attivatore, non è in grado di legare il suo interattore proteico. Si consideri che l'attività dei siti di regolazione dello splicing dipende anche dallo stato epigenetico della cromatina.

dei siti di splicing. Le modificazioni epigenetiche della cromatina (vedi cap. 17) hanno un ruolo molto importante nella regolazione del processo di splicing.

Uno dei sistemi più studiati di regolazione dello splicing riguarda la determinazione del sesso in *Drosophila melanogaster*.

In *Drosophila* il sesso è determinato dal rapporto tra i cromosomi X e gli autosomi, pari a 1 nella femmina e a 0,5 nel maschio. Il differente dosaggio del cromosoma X nei maschi e nelle femmine comporta un diverso livello di espressione di due geni codificanti i fattori trascrizionali *SisA* e *SisB*, localizzati sul cromosoma X, doppio nella femmina rispetto al maschio. Questi fattori trascrizionali attivano l'espressione del gene denominato *Sex-Lethal* (*Sxl*) a partire da un suo promotore precoce ($P_{e(early)}$). L'azione di questi attivatori trascrizionali è contrastata da un repressore trascrizionale codificato dal gene *Dpn* e localizzato sul cromosoma 2. Il gene *Sxl* risulterà espresso solo nelle femmine, dove gli attivatori espressi da *SisA* e *SisB* sono in eccesso rispetto al prodotto di *dpn*. Al contrario, nei maschi, in cui il livello di *SisA* e *SisB* è dimezzato, prevarrà l'effetto del repressore *Dpn* e *Sxl* non verrà espresso (fig. 13.19).

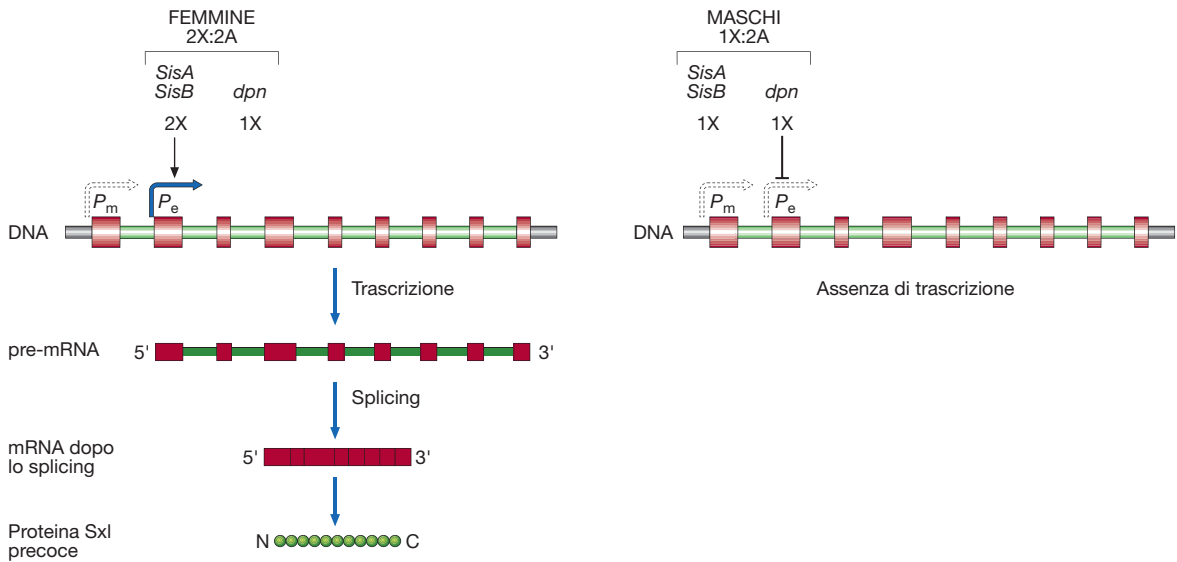


Figura 13.19 Rapporto tra cromosomi X e autosomi e determinazione del sesso in *Drosophila*. La trascrizione del gene *Sxl* di *Drosophila* a partire dal promotore precoce (P_e) ha luogo solo nelle

femmine, dove per un effetto di dosaggio tra cromosoma X e autosomi (femmine X:A = 2/2; maschi X:A = 1/2) gli attivatori SisA e SisB prevalgono sul repressore Dpn.

Sxl viene poi trascritto, sia nelle femmine che nei maschi, a partire da un secondo promotore denominato P_m (o di mantenimento) a monte di P_e . Il pre-mRNA così generato subirà un diverso processo di splicing nei maschi e nelle femmine (fig. 13.20), in quanto in queste ultime *Sxl*, già presente perché espressa a partire da P_e , è in grado di autoregolare lo splicing del suo pre-mRNA. *Sxl* funge da repressore dello splicing inducendo l'espressione di due isoforme di mRNA *Sxl*, di cui solo quella espressa nelle femmine è in grado di esprimere una proteina funzionale, che a sua volta controlla lo splicing del gene *tra* (transformer). Solo l'isoforma di *tra* espressa nelle femmine codifica per un prodotto funzionale che a differenza di *Sxl* funziona come attivatore dello splicing utilizzando come cofattore la proteina Tra2. In questo modo nei maschi e nelle femmine vengono prodotte due isoforme differenti della proteina double-sex (*dsx*). Queste due isoforme hanno una porzione N-terminale in comune e una porzione C-terminale maschio-specifica di 150 aa e femmina-specifica di 30 aa che inducono, rispettivamente, lo sviluppo dello stato maschile e femminile (fig. 13.20).

13.2 Editing

L'editing dell'RNA è un processo post-trascrizionale che genera molecole di RNA che differiscono dal loro stampo di DNA in una o più posizioni, modificando così l'informazione genetica presente nel gene. L'editing utilizza due diversi meccanismi:

- la **conversione** di una base in un'altra;
- l'**inserzione o delezione** di nucleotidi.

Gli effetti prodotti dall'editing vanno dalla sostituzione di un amminoacido in un altro, con conseguente alterazione delle proprietà strutturali e funzionali della proteina, alla creazione/eliminazione di codoni di inizio e terminazione, fino alla creazione di intere ORF, come nel caso dell'editing per inserzione/delezione che sarà descritto più avanti.

Questo processo riguarda solo siti specifici che vengono sottoposti a editing

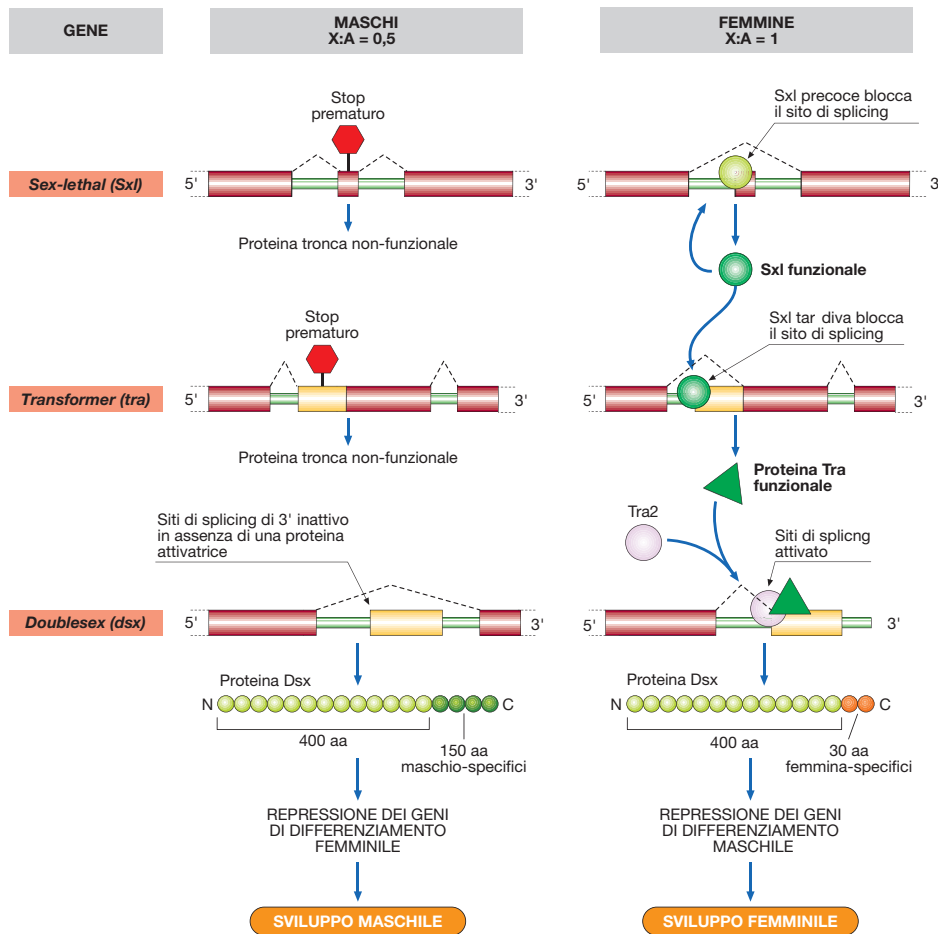


Figura 13.20 Determinazione del sesso in *Drosophila*. Azione dei repressori e attivatori dello splicing Sex-lethal (*Sxl*), Transformer (*Tra*) e Doublesex (*Dsx*) (vedi testo).

dall'azione enzimatica di fattori proteici o di RNP, costituite da fattori proteici e da molecole di RNA, che riconoscono motivi di sequenza e/o strutturali presenti in prossimità dei siti bersaglio, in un processo regolato in funzione dello stadio di sviluppo, del tessuto o della condizione fisiologica. La funzione svolta dall'editing è quella di apportare le opportune correzioni per rendere i trascritti funzionali o per modulare l'espressione di prodotti alternativi in diverse condizioni o tessuti. L'editing è un processo ubiquitariamente diffuso negli eucarioti e riguarda sia trascritti di geni nucleari che organellari. Modificazioni a carico dell'RNA imputabili al processo dell'editing sono state anche ritrovate in eucarioti unicellulari e alcuni virus.

Editing per conversione di basi

L'editing per conversione di basi è frequente negli organelli cellulari delle piante e, in particolare, nei mitocondri, in cui specifiche citosine sono deamminate in uracile ad opera di attività enzimatiche che coinvolgono proteine della famiglia PPR (caratterizzate da ripetizioni di 35 aa denominate "Pentatricopeptide Repeat"), che solo recentemente sono state trovate implicate nel fenomeno dell'editing dell'RNA. Nei cloroplasti e nei mitocondri di piante inferiori, inoltre, è stato osservato il **reverse editing** che comporta, quindi, la conversione inversa mediante amminazione dell'uracile in citosina. I fattori molecolari implicati nel reverse editing, tuttavia, non sono ancora noti con precisione.

In mitocondri e cloroplasti di piante è stato osservato come il fenomeno dell'editing sia fortemente gene- e organismo-specifico. Per esempio, il gene mitocondriale *matR* in *Arabidopsis thaliana* è editato mediante sostituzioni citosina-uracile in 14 posizioni, mentre lo stesso gene in *Beta vulgaris* e *Vitis vinifera* è modificato in 9 e 11 posizioni, rispettivamente. A differenza del gene *matR* in *Arabidopsis thaliana*, il gene *atp4* nella stessa pianta è editato in 9 posizioni e il gene *nad4* in ben 40 diversi siti.

Nonostante gli eventi di editing siano differenti in trascritti del medesimo gene in organismi differenti, il principale effetto dell'editing nei mitocondri di piante è quello di incrementare la similarità a livello proteico. In effetti, gran parte degli eventi a carico dei trascritti organellari di piante coinvolge la seconda posizione del codone, con conseguente sostituzione amminoacidica che generalmente ripristina un amminoacido evolutivamente conservato in altre specie.

Negli animali e, specialmente nei mammiferi, il fenomeno dell'editing avviene quasi esclusivamente per conversione di basi e non interessa il compartimento mitocondriale. Nelle cellule umane, e di altri pochi mammiferi in cui è stato investigato, l'editing può modificare specifiche basi mediante la conversione della citosina in uracile (C→U) o la deaminazione dell'adenosina in inosina (A→I, vedi fig. 2.2).

La conversione della citosina in uracile è un evento raro negli animali e l'esempio più noto è rappresentato dal gene per l'apolipoproteina B nei vertebrati, il cui trascritto, a seconda che subisca o meno il processo di editing, può codificare due proteine distinte ApoB100 e ApoB48, rispettivamente di 100 e 48 kDa, nel fegato e nell'intestino (fig. 13.21). La conversione della citosina in uracile nella posizione 6666 dell'mRNA, catalizzata dalla deaminasi APOBEC1, avviene solo nell'intestino e, comportando la trasformazione del codone CAA in UAA, introduce un codone di stop prematuro, responsabile della formazione dell'isoforma più corta di ApoB.

La conversione A→I (fig. 13.22A) è diffusa nei pre-mRNA dei mammiferi e soprattutto dell'uomo, in cui studi mirati mediante le tecnologie di sequenziamento massivo del trascrittoma hanno portato all'identificazione di migliaia di eventi

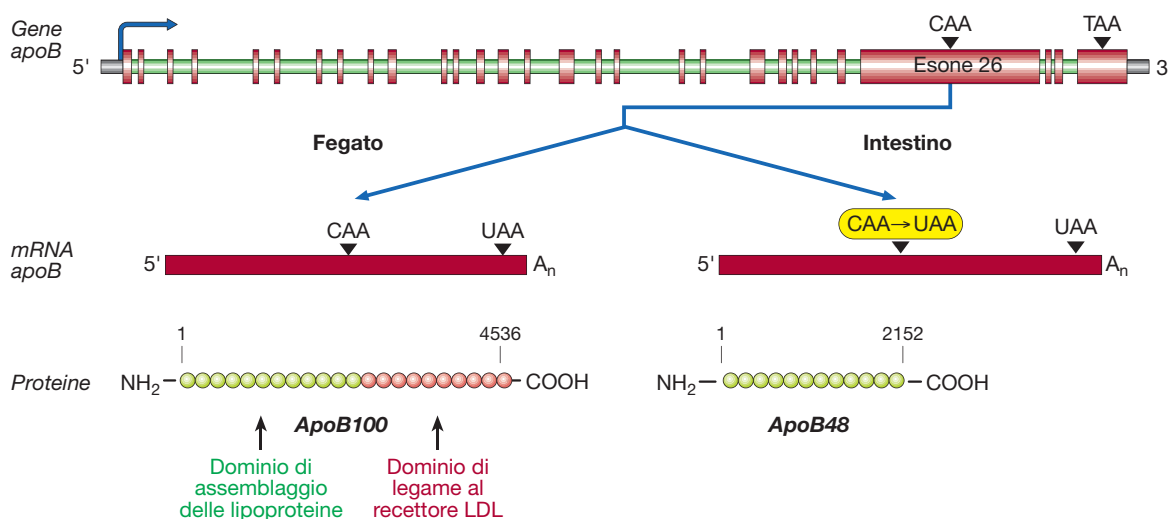


Figura 13.21 Editing C→U dell'mRNA per l'ApoB nei vertebrati. Il processo di editing, che genera un codone di stop prematuro, avviene in modo tessuto-specifico nell'intestino ma non nel fegato.

In questo modo viene prodotta la proteina ApoB100 nel fegato e ApoB48 nell'intestino.

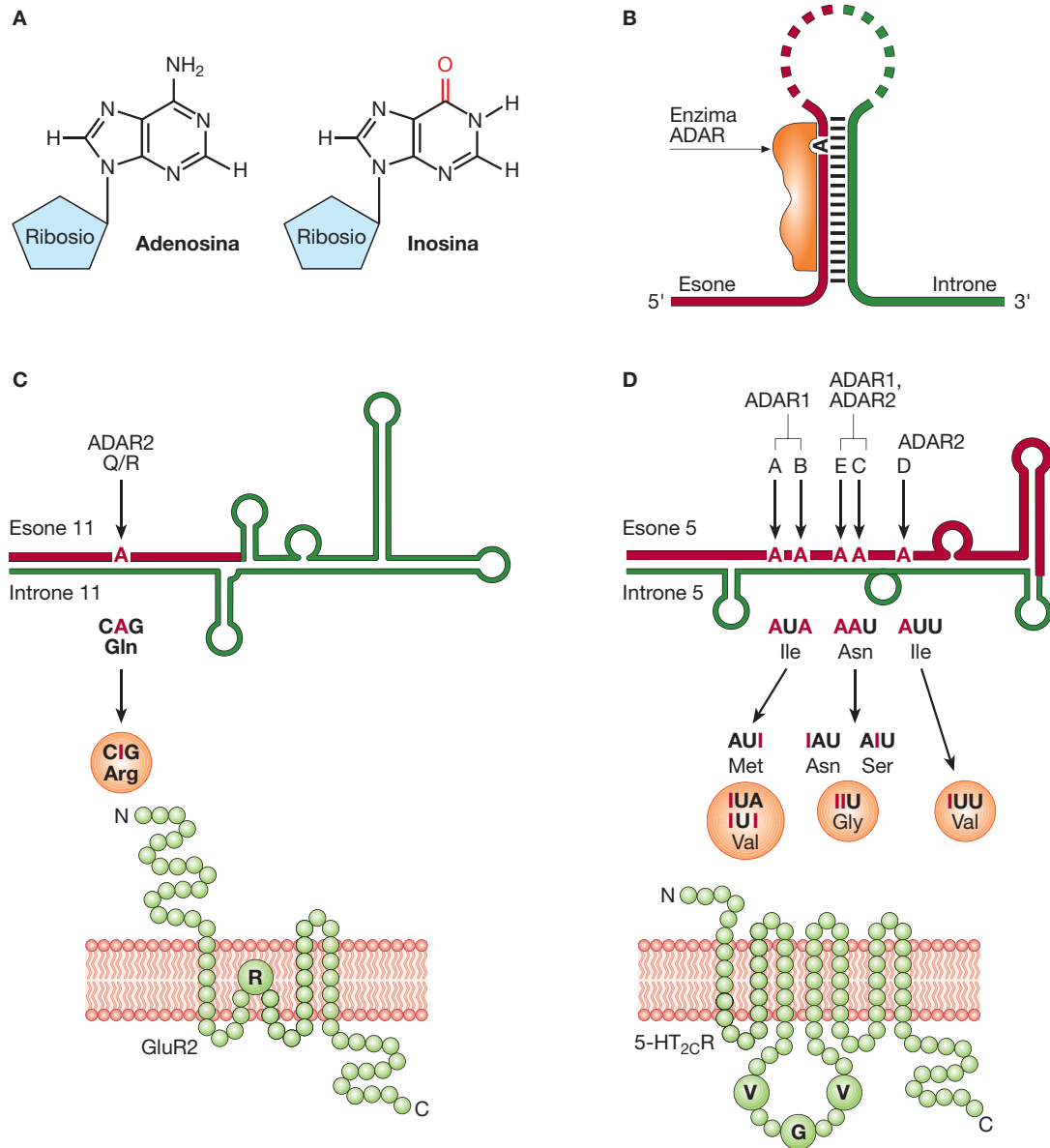


Figura 13.22 Conversione A→I. L'editing A→I nell'mRNA (A) comporta una reazione di deaminazione catalizzata da un enzima specifico (ADAR, *Adenosine Deaminase Acting on RNA*) che riconosce una regione di RNA a doppio filamento tra un esone (nel quale si trova il sito di editing) e il successivo introne (B). (C) L'editing A→I nell'mRNA per la subunità B del recettore del glutammato

comporta la sostituzione di una glutammina con arginina che altera le proprietà del recettore e la permeabilità della membrana al Ca^{2+} . (D) L'mRNA per il recettore della serotonina 5-HT_{2C}R subisce il processo di editing in cinque siti distinti che comportano tre sostituzioni amminoacidiche che alterano le proprietà della trasduzione del segnale che coinvolge le proteine G.

di editing a carico sia delle regioni codificanti proteine che delle porzioni non tradotte (UTR). Tuttavia, la maggior parte delle sostituzioni A→I osservate in cellule umane da diversi tessuti o condizioni sperimentali avviene in regioni ripetute e, particolarmente, in porzioni di RNA ricche di elementi *Alu*.

La deaminazione dell'adenosina in inosina richiede l'intervento di un enzima della famiglia ADAR (*Adenosine Deaminase Acting on RNA*) che riconosce come substrato RNA a doppio filamento formato dall'appaiamento di regioni complementari che possono essere un esone e l'introne a valle (fig. 13.22B) o regioni ripetute come due elementi *Alu* in orientamento opposto.

L'inosina generata dalle ADAR è normalmente riconosciuta e interpretata co-

me guanosina dai principali macchinari molecolari che agiscono sull'RNA, come l'apparato traduzionale o i diversi componenti dello spliceosoma. Il sequenziamento di un RNA sottoposto a editing A→I evidenzia quindi sostituzioni A→G che, a seconda della regione dell'RNA in cui avvengono, possono avere differenti ripercussioni funzionali. Difatti, la disregolazione dell'editing è stata associata a diverse patologie neurologiche umane come l'epilessia, la depressione, la schizofrenia o la sclerosi laterale amiotrofica. Attualmente, sembra che l'editing dell'RNA possa anche essere coinvolto nella abnorme proliferazione cellulare tipica dei tumori.

L'evento di editing A→I maggiormente caratterizzato è quello che interessa il gene per la subunità 2 del canale ionico AMPA che risponde al glutammato (GRIA2) in cui uno degli enzimi ADAR riconosce come substrato l'RNA a doppio filamento formato dall'appaiamento di un esone e del successivo introne (**fig. 13.22B**). L'editing in GRIA2 introduce cambiamenti nella sequenza amminoacidica. Per esempio, l'evento a carico di un codone CAG comporta la sostituzione di una glutammata (CAG) in arginina (CIG = CGG). Tale modificazione amminoacidica (**fig. 13.22C**) sembrerebbe modificare la permeabilità al calcio del canale AMPA, con potenziali conseguenze funzionali a carico principalmente del cervello e del sistema ematopoietico.

Altri eventi A→I particolarmente importanti per il corretto funzionamento del sistema nervoso centrale dei vertebrati sono quelli che interessano il recettore per la serotonina (**fig. 13.22D**). Alterazioni nello stato dell'editing di tale recettore, infatti, hanno mostrato una disregolazione nella neurotrasmissione tipica di patologie come la schizofrenia o la depressione grave.

Editing inserzionale

L'editing inserzionale, scoperto da Benne e collaboratori nel 1986, è stata la prima forma di editing a essere scoperta, nel protozoo *Trypanosoma brucei*, l'agente della malattia del sonno (**fig. 13.23A**). Questi protozoi sono caratterizzati dalla presenza di un singolo grande mitocondrio il cui genoma è organizzato in forma di un intricato network di molecole di DNA che includono circa 20-50 maxi-circoli e circa 5 000-10 000 mini-circoli.

I geni mitocondriali sono codificati nei maxi-circoli, ma producono RNA che diventano funzionali solo dopo che l'editing, che consiste nell'aggiunta e delezione di uridine in specifiche posizioni, sia stato completato (**fig. 13.23B e 13.23C**). L'editing può essere anche molto esteso, come nel caso del trascritto per la subunità 6 della ATPasi (A6) in cui vengono aggiunte e rimosse 447 e 28 uridine, rispettivamente, per creare un trascritto funzionale di 820 nt a partire da un precursore di 401 nt. I geni sono quindi presenti nel genoma in una forma criptica e sono difficilmente riconoscibili. Per questa ragione sono anche definiti **criptogeni**.

Le posizioni specifiche che dovranno essere editate e il numero corretto di U sono determinati dagli **RNA guida** (gRNA) codificati dai mini-circoli. Questi sono piccoli RNA che contengono una breve sequenza antisense che si appaia con il trascritto non editato a valle del sito in cui andranno inserite o eventualmente rimosse le uridine secondo il meccanismo schematicamente rappresentato in **figura 13.24**. Questo tipo di editing richiede un complesso enzimatico, detto **editosoma**, che riconosce la regione appaiata ed effettua l'editing attraverso l'azione di specifici enzimi, quali una endonucleasi, una terminal U transferasi (TUTasi) e una RNA ligasi. Il processo di editing inserzionale è generalmente molto esteso e richiede numerosi gRNA, secondo il meccanismo progressivo illustrato in **figura 13.25**, nel quale le modificazioni mediate dal primo gRNA consentono l'interazione e quindi le modificazioni mediate dal gRNA successivo, e così via.

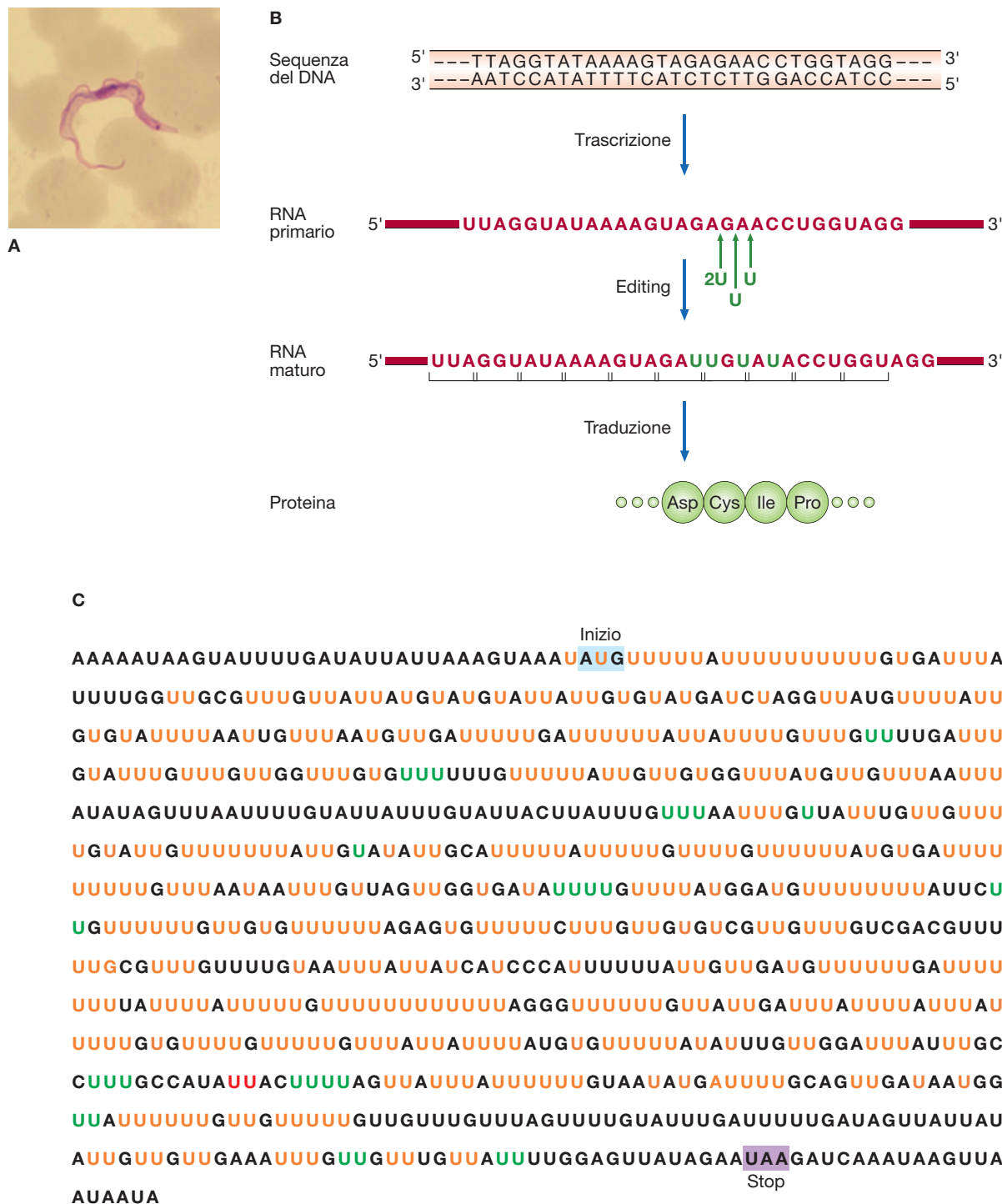


Figura 13.23 Editing inserzionale nel mitocondrio di *Trypanosoma*. (A) Microfotografia di *Trypanosoma brucei*, l'agente della malattia del sonno. (B) Gene *Cox II* (sunità 2 della citocromo ossidasi): è mostrato solo un breve tratto di questo gene il cui trascritto primario subisce un limitato editing che consiste nell'inserzione di 4 U che completano la sequenza dell'mRNA, il quale viene

poi tradotto in proteina. (C) Gene A6 (subunità 6 della ATPasi): il trascritto primario di questo gene subisce un editing molto esteso, come evidenziato nella sequenza dell'mRNA maturo qui mostrata, in cui i nucleotidi codificati a livello del DNA sono indicati in nero; il susseguente editing risulta nell'inserzione di numerose U (in rosso) e nella delezione di poche U (in verde).

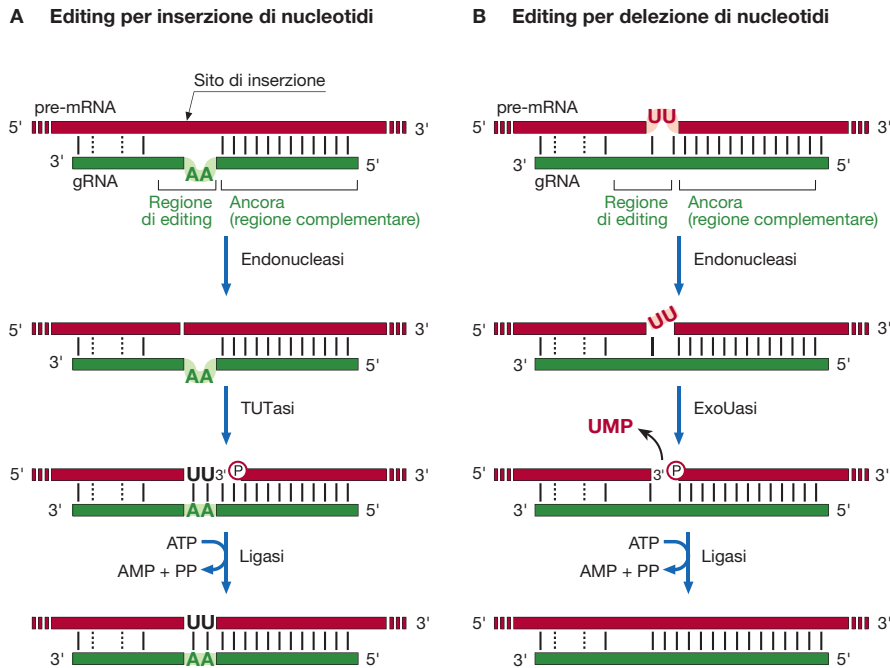


Figura 13.24 Meccanismo dell'editing inserzionale nel mitocondrio di *Trypanosoma*. Il meccanismo di editing comporta l'interazione tra il tratto di RNA che deve essere modificato (pre-mRNA) e uno specifico RNA guida (gRNA). Questo interagisce in prossimità del sito di editing per mezzo della sequenza ancora complementare al pre-mRNA per lo più nel tratto a valle. (A) Editing per inserzione di nucleotidi: adiacente al tratto di complementarietà si ha una regione (regione di editing) non appaiata per la presenza di A, due in questo caso, che non hanno le corrispettive U nel pre-mRNA. Una endonucleasi taglia in corrispondenza del sito di editing riconoscendo la regione non

appaiata con i gRNA e, successivamente, l'enzima TUTasi (*Terminal Uridil Transferasi*) aggiunge le U al 3'OH del frammento di pre-mRNA a monte. Una ligasi, infine, unisce le estremità 3'OH e 5'-fosfato ristabilendo la continuità della molecola di RNA. (B) Editing per delezione di nucleotidi: in questo caso la regione di editing non è appaiata per la mancanza di A in corrispondenza di 1 o più U, due in questo esempio, nel sito di editing del pre-mRNA. Anche in questo caso una endonucleasi taglia il pre-mRNA in corrispondenza del sito di editing e le U in eccesso vengono rimosse da una endonucleasi U specifica (ExoUasi); una ligasi, infine, ristabilisce la continuità della molecola di RNA.

Figura 13.25 Rappresentazione semplificata di un modello per un processo di editing inserzionale molto esteso. All'inizio la maggior parte dei gRNA non sono in grado di interagire con il pre-mRNA perché le loro regioni ancora non trovano nel pre-mRNA le sequenze complementari cui appaiarsi. Solo il primo gRNA (gRNA 1) si appaia mediante la sua sequenza ancora a una regione complementare che trova al 3' del pre-mRNA e guida l'editing della regione adiacente a monte. A questo punto il gRNA 2 può appaiarsi con la sequenza appena editata (i siti editati sono indicati da quadratini rosa), ora complementare alla sua propria sequenza ancora, e procedere all'editing della regione adiacente ulteriormente a monte. Così procedendo il processo di editing risale lungo il pre-mRNA dall'estremità 3' fino all'estremità 5'. L'ultimo gRNA (qui gRNA 6) potrebbe per esempio inserire una U mancante nel codone di inizio (AG che diventa così AUG), rendendo l'mRNA attivo nella traduzione.



13.3 Traslocazione nucleo-citoplasmatica degli RNA

Una volta completato il processo di maturazione all'interno del nucleo gli RNA devono essere traslocati nel citoplasma, dove potranno esercitare direttamente la loro funzione (ad es. mRNA) oppure, come nel caso di snRNA e snoRNA, costituire specifici complessi ribonucleoproteici per poi ritornare all'interno del nucleo.

Il movimento di ioni, metaboliti e altre piccole molecole attraverso la membrana nucleare avviene per diffusione passiva, mentre la traslocazione di molecole di dimensioni maggiori di 40 kDa richiede l'intervento di un apparato di traslocazione che regoli dinamicamente questo processo, controllando anche l'effettiva completa maturazione e integrità dei trascritti (**fig. 13.26**).

Le strutture deputate al trasporto attivo attraverso la membrana nucleare sono i **complessi dei pori nucleari (NPC, Nuclear Pore Complex)**, grandi complessi

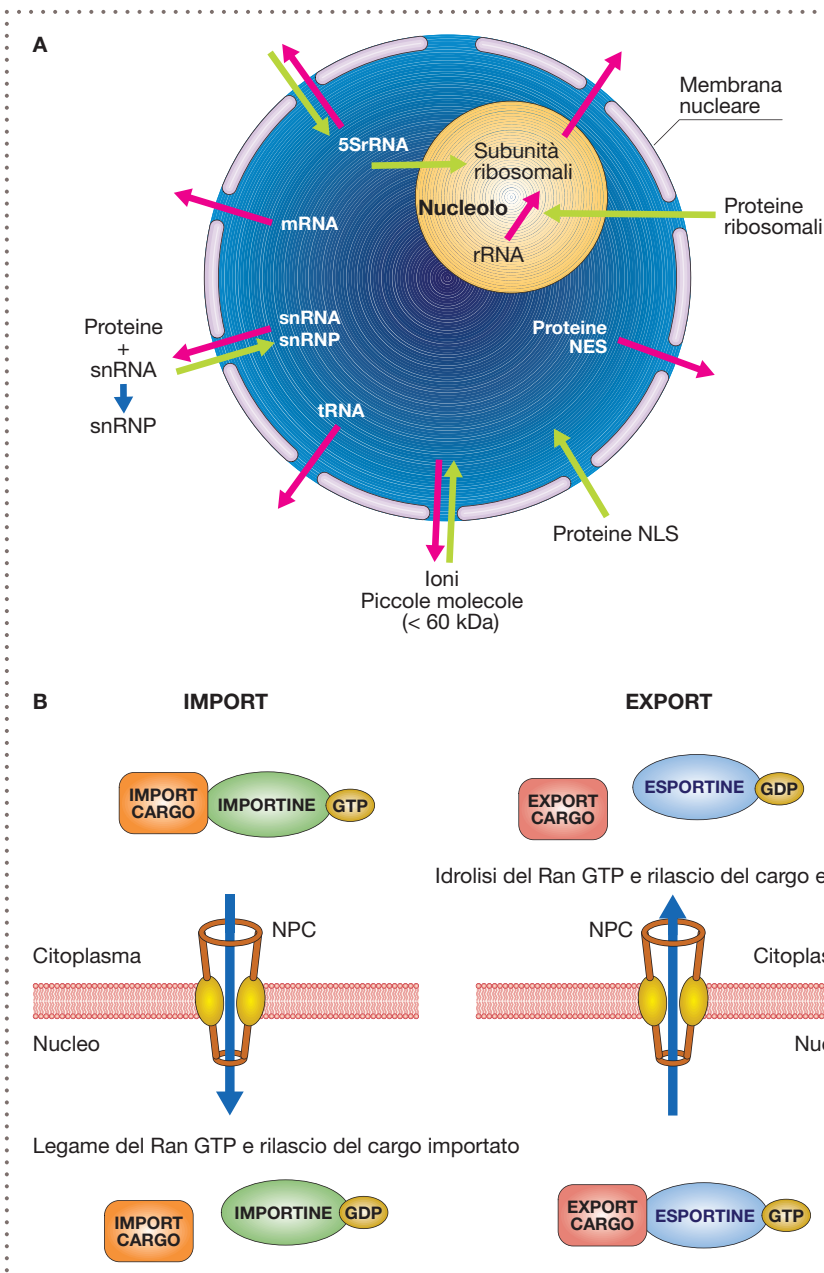


Figura 13.26 Movimento di ioni, metaboliti e altre piccole molecole attraverso la membrana nucleare. (A) Import (freccie in verde) ed export (freccie in rosso) di macromolecole attraverso la membrana nucleare e il nucleolo. (B) Il meccanismo attivo di traslocazione nucleo-citoplasmatica coinvolge i complessi dei pori nucleari (NPC, Nuclear Pore Complex) e proteine della classe delle carioferine (denominate importine o esportine a seconda della direzione del trasporto) cui si legano le macromolecole trasportate (dette anche cargo). Il processo richiede energia fornita dall'idrolisi di GTP da parte di una GTPasi denominata Ran.

multiproteici i cui componenti sono denominati **nucleoporine**. Gli NPC svolgono un ruolo essenziale nella crescita e nella progressione del ciclo cellulare, nella risposta a stimoli extracellulari e nello sviluppo. Oltre agli NPC il trasporto nucleo-citoplasmatico richiede altri fattori proteici in grado di riconoscere in modo specifico le macromolecole da trasportare, denominati **carioferine**, oppure importine o esportine in relazione alla direzionalità del processo. Le carioferine riconoscono specifici segnali denominati NLS (*Nuclear Localization Signal*) o NES (*Nuclear Export Signal*) sia a livello delle proteine che degli RNA, per poi mediarne la traslocazione, processo che richiede anche l'apporto energetico fornito dall'idrolisi di GTP catalizzata da una GTPasi denominata Ran.

I fattori proteici che agiscono come mediatori del trasporto attivo degli mRNA attraverso gli NPC controllano anche che il processo di maturazione sia stato completato e sia avvenuto correttamente. Per questo, via via che procede il processo di maturazione, sull'RNA si assemblano una serie di proteine specifiche per formare una ribonucleoproteina (RNP) che potrà essere indirizzata al processo di trasporto nucleo-citoplasmatico.

In particolare, gli mRNA formano mRNP che solo dopo un appropriato controllo di qualità vengono esportati nel citoplasma in un processo che richiede diversi fattori proteici, incluse proteine della famiglia SR, che abbiamo già incontrato nella regolazione dello splicing. Lo stato di fosforilazione di queste proteine, che normalmente vengono defosforilate in seguito al processo di splicing, è l'elemento discriminante per valutare l'integrità degli mRNP. Gli mRNP aberranti, invece, attivano la completa degradazione degli mRNA attraverso l'apparato dell'esosoma nucleare.

Lecture di approfondimento

- Black, D.L. (2003), «Mechanisms of alternative pre-messenger RNA splicing», *Annual Review of Biochemistry*, **72**, pp. 291-336.
- Brennicke, A., Marchfelder, A., Binder, S. (1999), «RNA editing», *FEMS Microbiology Reviews*, **23**(3), pp. 297-316.
- Fedor, M.J., Williamson, J.R. (2005), «The catalytic diversity of RNAs», *Nature Reviews Molecular Cell Biology*, **6**(5), pp. 399-412.
- Licatalosi, D.D., Darnell, R.B. (2010), «RNA processing and its regulation: global insights into biological networks», *Nature Reviews Genetics*, **11**(1), pp. 75-87.
- Wente, S.R. (2000), «Gatekeepers of the nucleus», *Science*, **288**(5470), pp. 1374-1377.



TRADUZIONE E SUE REGOLAZIONI

Come la maggior parte degli autori, consideriamo i termini “traduzione” e “sintesi proteica” sostanzialmente sinonimi, anche se sembrano enfatizzare due aspetti diversi del processo cui si riferiscono. Infatti, il termine “traduzione” richiama il concetto di decodificazione del messaggio genetico e appare quindi tipico di una visione genetica (informazionale) dei sistemi viventi, mentre il termine “sintesi proteica” richiama la polimerizzazione degli amminoacidi in proteine ed è quindi più consono a una visione biochimica (strutturale) della biologia.

La traduzione è uno dei processi più conservati fra tutti gli organismi viventi: è il processo cellulare più complesso per numero di componenti e di interazioni molecolari implicati e impegna una gran parte delle risorse energetiche della cellula. Infatti, l'apparato di sintesi proteica include, oltre all'mRNA e ai ribosomi (costituiti da 50-80 diverse proteine e vari RNA ribosomali), anche più di trenta tipi di tRNA (RNA di trasferimento), una ventina di enzimi implicati nell'attivazione degli amminoacidi, numerosi fattori proteici necessari per le fasi di inizio, allungamento e terminazione della traduzione. In una cellula procariotica ci sono circa 20000 ribosomi e 200000 molecole di tRNA per tradurre 1000-2000 molecole di mRNA; in complesso, tra il 30 e il 50% del peso secco di una cellula procariotica è dedicato alla sintesi proteica. Nelle cellule degli eucarioti superiori ribosomi, tRNA e mRNA sono molti di più, ma rappresentano una percentuale minore della massa secca essendo le cellule eucariotiche molto più grandi e strutturalmente molto più complesse delle cellule procariotiche.

Meccanicisticamente la traduzione consiste nella costruzione di proteine da parte del ribosoma, che, con l'ausilio di altri componenti dell'apparato di traduzione, scorre sull'mRNA in direzione 5'→3' decifrandone l'informazione codificata e catalizzando l'aggiunta progressiva di amminoacidi, uno alla volta, alla catena proteica nascente che viene sintetizzata sempre nella direzione ammino → carbossi-terminale. La velocità di allungamento della catena

amminoacidica è di circa 15 amminoacidi al secondo nei procarioti e 2 negli eucarioti. L'efficienza della produzione di proteine è aumentata dal fatto che più ribosomi possono legarsi in successione allo stesso mRNA e procedere in fila lungo di esso formando così un "polisoma" (contrazione di "poliribosoma"), cioè una molecola di mRNA con vari ribosomi associati, da 2 o 3 fino a una o due decine, a seconda della lunghezza della regione codificante dell'mRNA. Negli eucarioti i vari componenti dell'apparato di traduzione (mRNA, tRNA e ribosomi) vengono sintetizzati, maturati e assemblati nel nucleo, per poi essere esportati nel citoplasma, dove avviene la sintesi proteica, che quindi è un processo temporalmente distinto dalla trascrizione. Nei procarioti, al contrario, trascrizione e traduzione avvengono nello stesso comparto cellulare, cosicché un mRNA ancora in fase di trascrizione può iniziare a essere tradotto da ribosomi che si legano alla sua estremità 5'. Ciò è possibile perché la trascrizione procede in direzione 5'→3', che è anche la direzione di traduzione dell'mRNA da parte dei ribosomi, permettendo così un'interazione tra i due processi e, quindi, interessanti possibilità regolative.

Lo studio del meccanismo molecolare della sintesi proteica ha recentemente fatto un salto di qualità grazie alla straordinaria impresa scientifica e tecnologica che ha portato alla risoluzione a livello atomico della struttura del ribosoma. Ciò ha permesso infatti il passaggio da una descrizione alquanto schematica e superficiale delle strutture e interazioni molecolari coinvolte nella traduzione a un loro studio e comprensione a livello molecolare molto più fine, compito che terrà impegnati molti gruppi di ricerca nei prossimi anni.

In questa parte del volume prima descriveremo i principali componenti dell'apparato di traduzione (ribosomi, tRNA, mRNA) per passare poi alla descrizione del meccanismo distinguendo per convenienza le tre fasi, di inizio, allungamento e terminazione, sempre confrontando i processi nei procarioti e negli eucarioti. Sarà infine affrontato il problema della regolazione della traduzione a livello globale e a livello di specifici mRNA.

14

Apparato di traduzione e suoi componenti

14.1 Ribosoma

14.2 RNA transfer (tRNA, RNA di trasferimento) e aminoacil-tRNA sintetasi

14.3 RNA messaggero (mRNA)

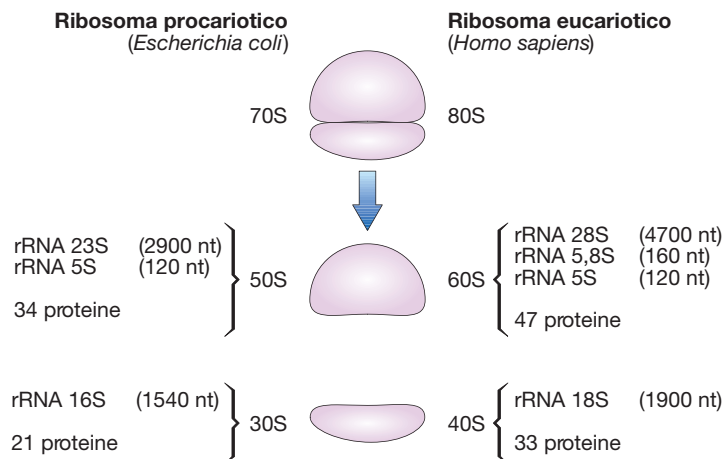
La sintesi proteica coinvolge interazioni fra tre tipi di molecole di RNA: l'mRNA (RNA messaggero), il tRNA ("transfer RNA", RNA di trasferimento) e l'rRNA (RNA ribosomale). Questi vengono tutti prodotti mediante il processo di "trascrizione" del DNA da parte delle RNA polimerasi e vengono successivamente "maturati" mediante una serie di importanti processi di modificazione. Oltre a questi RNA, sono necessari vari fattori proteici per le fasi di inizio, di allungamento e di terminazione della traduzione. Infine, la sintesi proteica necessita di energia per far funzionare il complesso macchinario di decifrazione del codice e per la formazione dei legami peptidici tra gli amminoacidi. Tale energia è fornita dall'idrolisi di GTP e di ATP. Descriveremo qui struttura e ruolo dei principali componenti dell'apparato di traduzione senza entrare nel merito dei processi della loro sintesi, maturazione e assemblaggio, che saranno invece trattati in altre parti di questo testo.

14.1 Ribosoma

Già dagli anni 1940-1950 era stata osservata una stretta correlazione tra la quantità di RNA nelle diverse cellule e la capacità di queste ultime di sintesi proteica, il che suggeriva un ruolo diretto dell'RNA nella sintesi proteica. Negli anni successivi, l'introduzione dell'uso di radioisotopi (in particolare ^{14}C e ^3H) per studiare l'incorporazione di amminoacidi nelle proteine di nuova sintesi e l'uso di metodi di microscopia elettronica e di ultracentrifugazione hanno permesso di identificare piccole particelle dense contenenti RNA, in parte libere nel citoplasma e in parte attaccate al reticolo endoplasmatico, che venivano indicate con vari nomi quali granuli basofili, microsomi, nucleoproteine e altri. Divenne presto chiaro che questo componente particolare rappresentava la sede della sintesi proteica. A partire dal 1960 comparve nella letteratura il termine "ribosoma".

I **ribosomi** sono delle "macchine" molecolari che sintetizzano le proteine decodificando l'informazione portata dall'mRNA. Nonostante alcune differenze di grandezza e composizione che si riscontrano tra ribosomi di vari organismi, la loro struttura e funzione sono rimaste altamente conservate durante l'evoluzione. Tutti i ribosomi procariotici ed eucariotici sono delle particelle ribonucleoproteiche, cioè sono composti da alcune molecole di RNA (**rRNA = RNA ribosomale**) e da numerose proteine (**rp = r-proteine = proteine ribosomali**). rRNA e r-proteine

Figura 14.1 Composizione molecolare dei ribosomi procariotici (*Escherichia coli*) ed eucariotici (*Homo sapiens*). I valori di lunghezza degli rRNA sono arrotondati; il numero di r-proteine presenta un piccolo margine di incertezza. Sia nei procarioti che negli eucarioti c'è una certa variabilità tra specie diverse.



sono assemblati a costituire due subunità ribosomali distinte, una maggiore e una minore. I ribosomi, le loro subunità e gli RNA che li compongono vengono generalmente denominati in base alla loro velocità di sedimentazione, misurata mediante ultracentrifugazione ed espressa in unità Svedberg (S) (vedi par. 20.3). Così i ribosomi procariotici vengono chiamati 70S perché hanno una velocità di sedimentazione di 70S, mentre quelli citoplasmatici eucariotici, che sono un po' più grandi e sedimentano più velocemente, vengono chiamati 80S. Come schematizzato in **figura 14.1**, i ribosomi procariotici 70S sono costituiti dalle due subunità chiamate 30S e 50S e quelli eucariotici 80S dalle subunità 40S e 60S. Va notato che i valori in S delle subunità non sono additivi rispetto a quelli dei ribosomi interi perché la velocità di sedimentazione non è direttamente proporzionale alla massa. Inoltre occorre precisare che questi valori di sedimentazione dei ribosomi, delle loro subunità e degli RNA ribosomali, sono valori nominali, utili per la loro identificazione ma molto approssimativi. La **figura 14.1** mostra anche come ciascuna subunità sia costituita da una o più molecole di rRNA (RNA ribosomale) e da numerose r-proteine (proteine ribosomali). La subunità 30S dei ribosomi procariotici contiene l'rRNA 16S costituito da circa 1540 nt (nt = nucleotidi) e 21 r-proteine diverse, mentre la subunità 50S contiene l'rRNA 23S (2900 nt), il piccolo RNA ribosomale 5S (120 nt) e 34 r-proteine. Nel complesso la massa del ribosoma procariotico è di circa 2500 kDa, ha un diametro di circa 250 Å ed è costituito per circa metà da RNA e metà da proteine. Il ribosoma eucariotico è simile a quello procariotico, ma ha rRNA più lunghi per l'inserzione, in regioni specifiche, di sequenze chiamate **sequenze di espansione**; ha anche, nella subunità maggiore, un piccolo rRNA in più, l'rRNA 5,8S (160 nt), che in realtà corrisponde all'estremità 5' dell'rRNA 23S batterico. Inoltre il ribosoma eucariotico ha, rispetto a quello procariotico, un numero maggiore di proteine, che nei vertebrati sono circa 80, e di conseguenza anche una massa maggiore, circa 4200 kDa. Va detto che le cellule eucariotiche contengono, oltre ai tipici ribosomi 80S citoplasmatici, anche ribosomi all'interno dei mitocondri e nel caso delle piante anche nei cloroplasti. Questi ribosomi sono per vari aspetti più simili ai 70S procariotici che non agli 80S eucariotici, a ricordarci l'origine di questi organelli come endosimbionti procariotici.

Il numero di ribosomi per cellula è molto variabile a seconda del tipo di cellula e delle condizioni di crescita. Una cellula di *Escherichia coli* contiene circa 20 000 ribosomi, mentre una cellula eucariotica ne contiene da 100 000 a un milione, ma cellule particolarmente grandi, come per esempio gli oociti, ne possono contenere anche miliardi.

Capire a fondo il meccanismo di sintesi proteica comporta la conoscenza

dettagliata della struttura dei ribosomi. Quattro decenni di lavoro di numerosi gruppi di ricerca hanno portato nel 2001 alla risoluzione a livello atomico della struttura del ribosoma 70S. L'obiettivo è stato raggiunto utilizzando, nel tempo, vari approcci e tecniche progressivamente più complesse e raffinate.

rRNA

Le prime informazioni strutturali hanno riguardato gli rRNA. Questi sono i primi RNA a essere stati purificati e analizzati grazie soprattutto alla loro abbondanza: rappresentano infatti circa l'80% dell'RNA delle cellule. L'analisi della sequenza nucleotidica degli rRNA di molte specie ha messo in evidenza che essi hanno una struttura secondaria dovuta ad appaiamenti di basi intramolecolari tra regioni complementari, come mostrato in **figura 14.2**. Le loro sequenze nucleotidiche, e soprattutto le loro strutture secondarie, sono evolutivamente molto conservate. Le regioni appaiate assumono la classica struttura a doppia elica e nel loro insieme si ripiegano, nel ribosoma, nella complessa struttura tridimensionale mostrata in **figura 14.3**, che non si discosta molto dalla forma che avrà la subunità ribosomale matura con le sue r-proteine. Il fatto che la struttura secondaria degli rRNA sia molto conservata, più di quanto non lo sia la sequenza nucleotidica, sta a indicare che quello che è funzionalmente rilevante, e quindi sotto pressione selettiva, è la struttura secondaria con la formazione delle numerose forcine, e il ripiegamento di queste nei vari domini della struttura tridimensionale. Se in termini e tempi evolutivi una mutazione cambia la sequenza ed elimina un appaiamento di basi, questo può essere ristabilito ripristinando la situazione iniziale con una retromutazione, ma anche mutando il nucleotide sull'altro filamento e ristabilendo una complementarità di basi, anche se diversa dall'originale. Un'evoluzione di questo genere fa sì che, nei tempi evolutivi, le sequenze nucleotidiche cambino progressivamente seppure molto lentamente, mantenendo però sostanzialmente invariata la struttura secondaria.

Gli rRNA, come anche i tRNA e altre classi di RNA, contengono oltre ai quattro tipi canonici di nucleotidi, un certo numero di nucleotidi insoliti (**fig. 14.4**). Questi non sono inseriti negli RNA al momento della sua trascrizione ma sono il prodotto di modificazioni enzimatiche post-trascrizionali di nucleotidi normali che si trovano in specifiche posizioni della molecola di RNA. Nel caso degli rRNA le modificazioni più frequenti che si riscontrano, ma non le uniche, sono l'isomerizzazione di alcune uridine in pseudouridine e la metilazione della posizione 2' del ribosio di alcuni nucleotidi. Queste modificazioni sono a carico di specifici nucleotidi che si trovano per lo più in regioni evolutivamente conservate. Negli eucarioti queste modificazioni sono abbastanza frequenti (per esempio ci sono un centinaio di pseudouridine e un centinaio di ribosi metilati nell'rRNA umano), mentre nei procarioti sono meno frequenti. I meccanismi responsabili di queste modificazioni chimiche delle basi sono descritti nel capitolo 12 sulla maturazione e sul processamento degli RNA.

r-proteine

Le proteine ribosomali sono in genere proteine non molto grandi, quasi tutte basiche e per lo più evolutivamente molto conservate. Mentre alcune r-proteine sono ancora ben riconoscibili nel confronto procarioti/eucarioti, in questi ultimi si osserva la comparsa di nuove proteine non presenti nei procarioti. Le r-proteine vengono identificate come rpS1, rpS2, rpS3 ecc. per la subunità minore e rpL1, rpL2, rpL3 ecc. per la subunità maggiore, dove "rp" sta per *ribosomal protein* mentre S e L indicano la subunità grande *large* e la subunità piccola *small*.

Informazioni sul tipo di interazioni tra le r-proteine e gli rRNA sono state ottenute con vari approcci sperimentali negli anni sessanta e settanta. Trattando

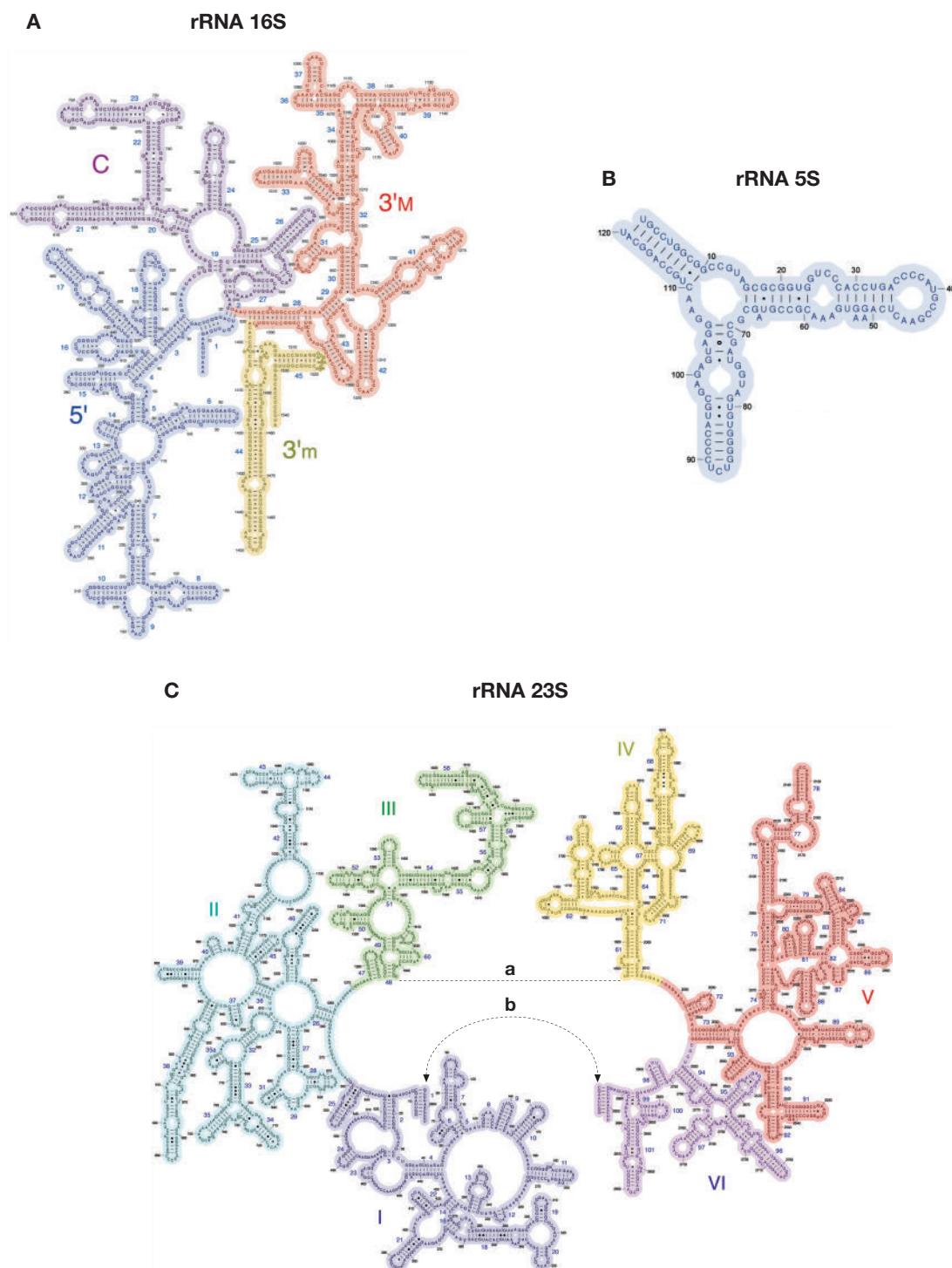
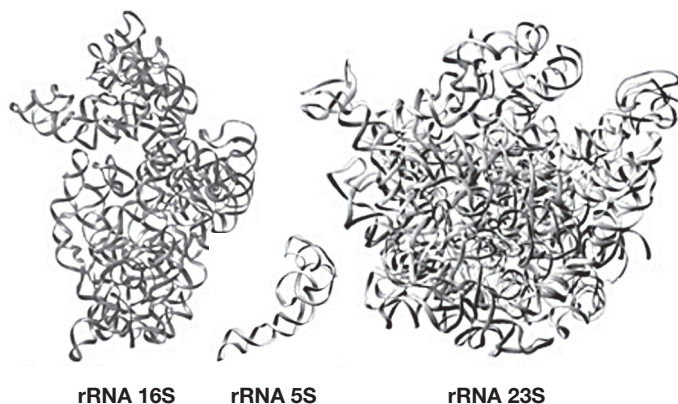


Figura 14.2 Strutture secondarie degli rRNA 16S, 5S e 23S di *Escherichia coli*. Per esigenze grafiche nel pannello (C) la struttura dell'rRNA 23S è stata divisa in due parti: la discontinuità è indicata dalla linea a. La doppia freccia b indica le due estremità 5' e 3' della molecola, che sono in realtà appaiate tra loro. Queste strutture sono state determinate e confermate con diversi approcci sperimentali. Innanzitutto, partendo dalle sequenze nucleotidiche degli RNA, si è cercato di ottimizzare gli appaiamenti intramolecolari di basi. Il confronto delle sequenze di

specie diverse ha permesso poi di confermare gli appaiamenti; infatti rRNA provenienti da specie diverse hanno strutture secondarie molto simili, pur con sostanziali differenze a livello della sequenza. Infine, dati di cristallografia hanno contribuito alla soluzione del problema. Le strutture secondarie degli rRNA 16S e 23S sono suddivise in vari domini (diversi colori), 4 nel 16S e 6 nel 23S, definiti da appaiamenti di basi a lunga distanza. (Fonte: Center for Molecular Biology of RNA, University of California, Santa Cruz, http://rna.ucsc.edu/rnacenter/noller_lab.html.)

Figura 14.3 Strutture tridimensionali degli rRNA 16S, 5S e 23S di *Escherichia coli*. Le sequenze appaiate nella struttura secondaria di figura 14.2 formano tratti di doppie eliche che, ulteriormente ripiegati, risultano nella conformazione qui rappresentata, che è molto simile a quella che si riscontra poi nelle subunità ribosomali. (Fonte: Center for Molecular Biology of RNA, University of California, Santa Cruz, http://rna.ucsc.edu/rnacenter/noller_lab.html.)

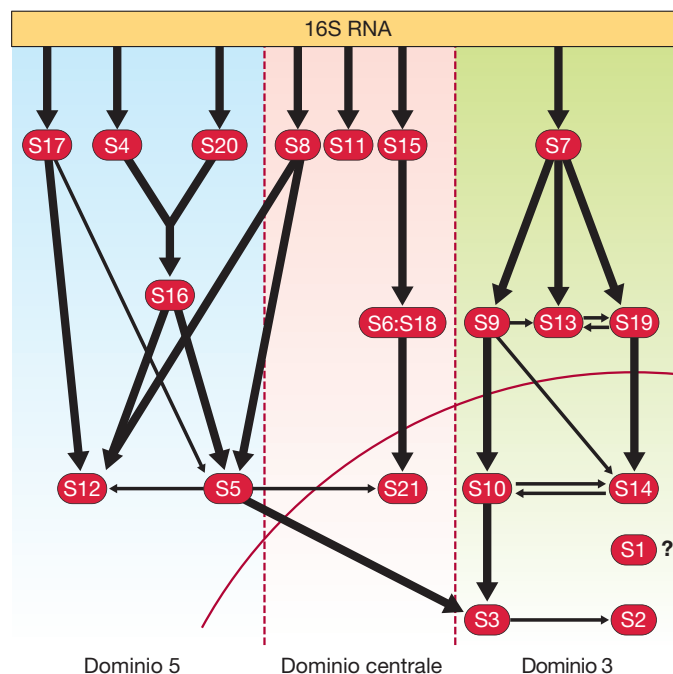


Base normale	Base modificata	Modificazione
Uridina 	Pseudouridina (Ψ) 	Si origina per isomerizzazione dell'uridina: l'anello aromatico è legato allo zucchero tramite il carbonio in posizione 5 anziché l'azoto in posizione 1
	Diidrouridina (D) 	Deriva dall'uridina per saturazione del doppio legame tra i carboni 5 e 6
	Tio-uridina (S ⁴ U) 	Deriva dall'uridina per sostituzione dell'ossigeno in posizione 2 con uno zolfo
Adenosina 	Inosina (I) 	Si origina per deaminazione di una adenosina
Guanosina 	7-Metilguanosina (m ⁷ G) 	Si origina per metilazione di una guanosina

Figura 14.4 Esempi di modificazioni chimiche di alcune basi che si trovano negli rRNA e nei tRNA. In rosso sono indicate le modificazioni rispetto alle basi canoniche.

subunità ribosomali batteriche con soluzioni ad alto contenuto salino (in genere CsCl o LiCl) e in assenza di ioni Mg^{2+} , le r-proteine possono essere disassemblate dagli rRNA ed essere poi separate l'una dall'altra mediante metodi cromatografici. È stato osservato che questo disassemblaggio avviene in maniera discontinua con gruppi di r-proteine che si dissociano in successione mano a mano che si alza la concentrazione salina. Questo disassemblaggio può essere invertito riabbassando la concentrazione salina e fornendo Mg^{2+} . Esattamente allo stesso modo si possono ricostituire subunità ribosomali funzionanti, cioè capaci di sostenere sintesi proteica *in vitro*, partendo da tutti i singoli componenti, rRNA e r-proteine, precedentemente purificati. Affinché la ricostituzione del ribosoma avvenga in maniera corretta, sono necessari vari accorgimenti che dimostrano la complessità del sistema. Innanzitutto le r-proteine vanno aggiunte all'rRNA in gruppi con un ordine definito che, non a caso, è lo stesso ordine, ma invertito,

Figura 14.5 Mappa di assemblaggio della subunità ribosomale 30S di *Escherichia coli*. Lo schema mostra quali proteine si legano per prime direttamente ai tre domini dell'rRNA 16S e quali si aggiungono successivamente in dipendenza di altre legate precedentemente. Le frecce nere indicano l'effetto di una proteina sul legame di una o più altre proteine. La linea rossa separa le proteine che si trovano legate nell'intermedio di ricostituzione della subunità da quelle che non sono presenti nell'intermedio. [Modificata da: Talkington, M.W., Siuzdac, G., Williamson, J.R. (2005), *Nature*, 438, pp. 628-632.]



con cui si sono dissociate dal ribosoma. Per esempio, come mostrato in **figura 14.5**, il primo passaggio della ricostituzione della subunità 30S del ribosoma batterico è l'associazione dell'rRNA 16S con un gruppo di 15 r-proteine, che sono le ultime che si erano staccate al momento della dissociazione della subunità ribosomale. Solo successivamente è possibile ottenere l'assemblaggio delle rimanenti 6 r-proteine, che sono quelle che per prime si erano dissociate. È stato osservato che all'interno di questi gruppi anche le singole r-proteine si associano alla costituenda ribonucleoproteina con un ordine determinato. Alcune si associano per prime legandosi direttamente all'rRNA, altre si possono associare solo dopo che già si sono assemblate le prime. Alcuni passaggi della riassociazione avvengono anche a freddo (0 °C), mentre altri richiedono un innalzamento della temperatura (40 °C), indicando che è coinvolto un cambiamento conformazionale della ribonucleoproteina che si va ricostituendo. Va infine detto che, quando è stata studiata la produzione di nuovi ribosomi *in vivo* nella cellula di *E. coli*, si è trovato che l'ordine di assemblaggio delle r-proteine sull'rRNA è molto simile a quello che è stato dimostrato *in vitro*.

Altre importanti informazioni sulle interazioni tra r-proteine e rRNA nella struttura del ribosoma sono state ottenute individuando e analizzando interazioni tra singole r-proteine e frammenti di rRNA o mediante induzione di legami crociati (*cross-linking*) per individuare r-proteine adiacenti nella struttura del ribosoma.

Struttura tridimensionale

Informazioni sulla struttura tridimensionale del ribosoma vennero inizialmente ottenute mediante tecniche di microscopia elettronica. Infatti, le subunità ribosomali sono abbastanza grandi da poter essere osservate, seppure a bassa risoluzione, al microscopio elettronico. Come si può vedere in **figura 14.6**, le due subunità ribosomali non si presentano amorfe, ma appaiono avere una struttura ben definita. Al problema della bassa risoluzione si è in parte ovviato sovrapponendo centinaia di immagini della stessa subunità e con lo stesso orientamento.

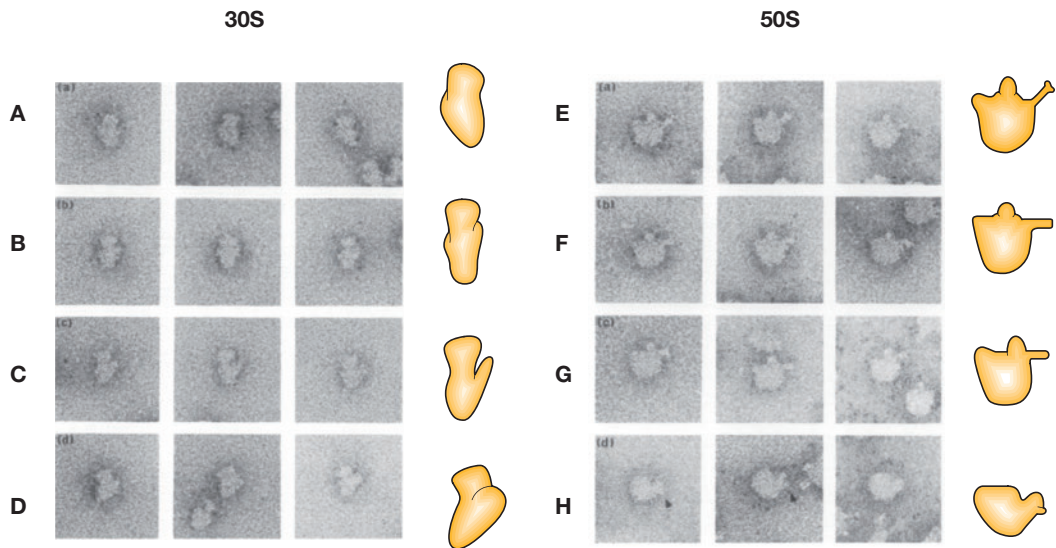
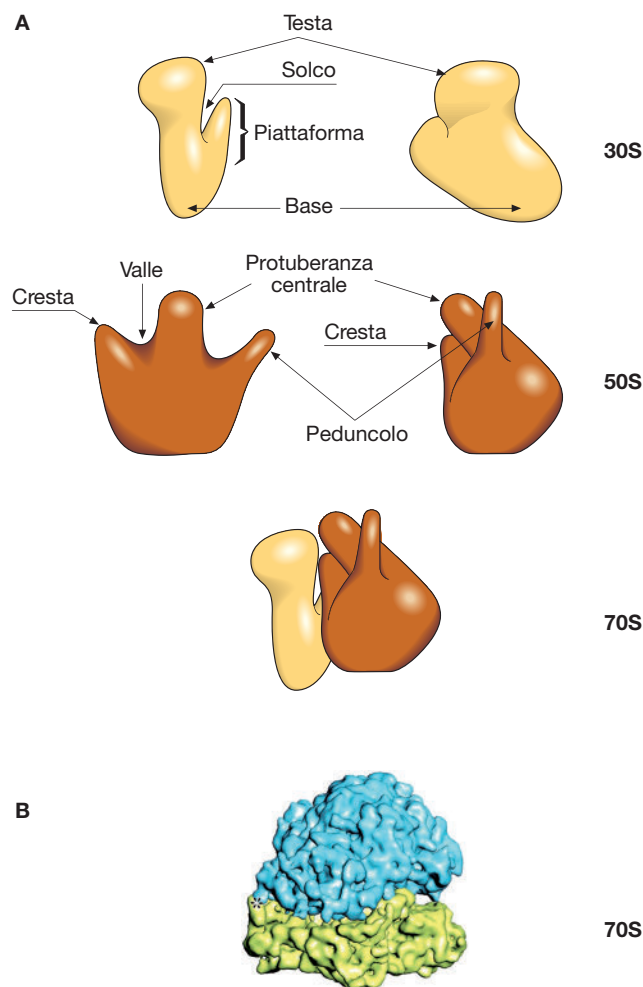


Figura 14.6 Fotografie al microscopio elettronico delle subunità ribosomali 30S e 50S di *Escherichia coli*. Le immagini delle subunità 30S (pannelli A-D) e 50S (E-H) sono in quattro orientamenti diversi. Per ogni orientamento sono mostrate tre fotografie e, a fianco, l'immagine derivata dall'osservazione di numerose fotografie. [Modificata da: Lake, J.A. (1976), *J. Mol. Biol.*, 105, pp. 131-159.]

Questo artificio ha permesso di costruire i primi modelli tridimensionali delle due subunità ribosomali 50S e 30S e, successivamente, dell'intero ribosoma 70S. Nel frattempo, altre informazioni strutturali venivano acquisite mediante approcci e tecniche diverse, tra cui la localizzazione delle r-proteine nella struttura del ribosoma mediante l'uso di anticorpi specifici per alcune di esse e l'analisi di interazioni tra singole r-proteine e specifiche regioni della struttura secondaria degli rRNA. Negli anni settanta, vari gruppi di ricerca, tra cui quelli diretti da James A. Lake, da Georg Stöffler e da Alexander S. Spirin, hanno utilizzato queste tecniche sia su subunità intere che su frammenti, giungendo a proporre modelli abbastanza realistici della struttura del ribosoma 70S, modelli che sono andati raffinandosi col tempo. La **figura 14.7A** mostra un esempio dei modelli che, negli anni ottanta, venivano proposti per le due subunità e per l'intero ribosoma 70S di *E. coli*. Nella struttura della subunità minore 30S distinguiamo una "base" (o "corpo") e una "testa"; dalla base protrude una "piattaforma" che è separata dalla testa da un "solco". Nella subunità maggiore 50S osserviamo una "protuberanza centrale" fiancheggiata da un lato da un "peduncolo" e dall'altro da una "cresta". Quest'ultima è separata dalla protuberanza centrale da una "valle". Successivamente, l'ulteriore affinamento delle tecniche di microscopia elettronica ha portato, negli anni novanta, a modelli del tipo di quello mostrato in **figura 14.7B**.

Tuttavia, la definizione ad alta risoluzione della struttura tridimensionale del ribosoma procariotico è stata raggiunta solo nel 2000-2001 con la tecnica di diffrazione dei raggi X. Data la grandezza e complessità del ribosoma, questa può essere considerata una delle più grandi imprese scientifiche mai affrontate e ha visto coinvolti vari grandi gruppi di ricerca, tra cui quelli diretti da Harry F. Noller, Venkatraman Ramakrishnan, Thomas A. Steitz e Ada Yonath (**fig. 14.8**). È interessante notare che questi importanti studi strutturali sono inizialmente stati fatti su ribosomi preparati da organismi estremofili, quali *Thermus thermophilus* e *Haloarcula marismortui*, che proprio perché crescono, rispettivamente, in condizioni di alta temperatura e di alta concentrazione salina, hanno ribosomi

Figura 14.7 Modelli strutturali del ribosoma. (A) Modelli di fine anni settanta delle subunità 30S e 50S e dell'intero ribosoma 70S di *Escherichia coli* visti da diverse prospettive. [Modificata da: Lake, J.A. (1995), *Annu. Rev. Biochem.*, 54, pp. 504-530]. (B) Modello a maggiore risoluzione proposto negli anni duemila per l'intero ribosoma 70S del ribosoma di *Trypanosoma cruzi*. [Fonte: Gabashvili, I.S. et al. (2000), *Cell*, 100, pp. 537-549.]



con struttura particolarmente stabile e, quindi, adatti a sopportare le condizioni non ottimali a cui vengono sottoposti durante l'analisi. Va anche detto che, a causa della grandezza e complessità della struttura, le immagini di diffrazione dei raggi X prodotte da cristalli di ribosomi non sono di per sé interpretabili. L'interpretazione è stata possibile perché, grazie alle tecniche di cui si è detto sopra, si avevano già molte informazioni e un modello abbastanza corretto, seppure approssimato, della struttura del ribosoma.

La **figura 14.9** riporta la struttura delle subunità 30S e 50S del ribosoma del batterio *Thermus thermophilus*. Il confronto con la **figura 14.3**, che riportava la struttura tridimensionale degli rRNA purificati, mostra che c'è una notevole somiglianza strutturale tra gli rRNA e le corrispondenti subunità ribosomali, indicando che la struttura complessiva del ribosoma è dovuta soprattutto agli rRNA, che con la loro struttura tridimensionale costituiscono un'impalcatura su cui si assemblano le r-proteine; come abbiamo già avuto modo di dire, alcune di queste stabiliscono legami direttamente con l'rRNA mentre altre vi si legano solo dopo che si sono legate le prime. Le r-proteine si trovano per lo più nella parte esterna del ribosoma e hanno in genere struttura globulare, spesso con prolungamenti che si infilano nella struttura dell'rRNA.

Questa conoscenza della struttura ad alta risoluzione del ribosoma ha permesso di iniziare studi a livello molecolare molto fine delle interazioni delle due subunità e degli altri componenti dell'apparato di traduzione durante il processo



A



B



C



D

Figura 14.8 La risoluzione della struttura del ribosoma ha visto coinvolti numerosi gruppi di ricerca; tra questi, soprattutto quelli diretti da Harry F. Noller (University of California a Santa Cruz) (A), Ada Yonath (Weizmann Institute of Science, Israele) (B), Thomas A. Steitz (Yale University, New Haven) (C) e Venki Ramakrishnan (MRC Laboratory of Molecular Biology, Cambridge, England) (D).

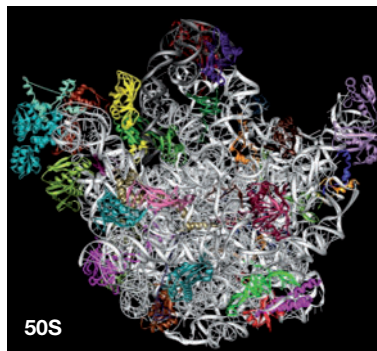
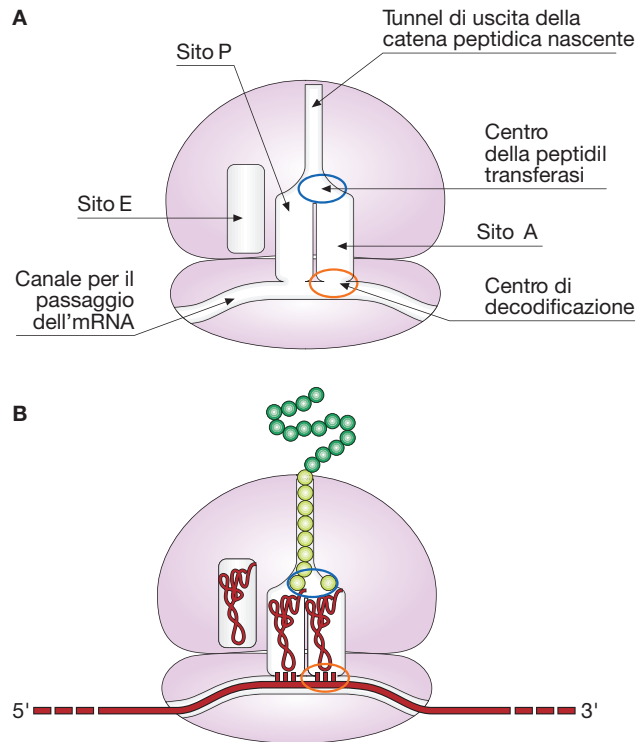


Figura 14.9 Struttura ad alta risoluzione delle subunità ribosomali 30S e 50S di *Thermus thermophilus*. La struttura è stata ottenuta mediante la tecnica di diffrazione dei raggi X di cristalli di subunità ribosomali, complementata da dati ottenuti con varie altre tecniche analitiche. Il nastro grigio rappresenta l'RNA. Le proteine sono rappresentate in colori vari. Per ambedue le subunità è mostrato il lato esterno rispetto alla struttura del ribosoma 70S. (Fonte: www.mzc-lmb.com.ac.uk/ribo.)

di sintesi proteica. Tuttavia, per descrivere qui i siti e centri attivi del ribosoma e le sue interazioni funzionali è conveniente tornare a un'immagine schematizzata e semplificata. Come mostrato in **figura 14.10A**, il ribosoma ha tre siti di legame per il tRNA: il **sito A** (Amminoacilico o Accettore), che lega il tRNA amminoacilato in ingresso, il **sito P** (Peptidilico), che lega l'ultimo tRNA entrato e porta la catena peptidica nascente, e il **sito E** (Exit, uscita), che lega il tRNA ormai scarico che deve essere rilasciato. I siti A e P si trovano a cavallo tra le due

Figura 14.10 Struttura del ribosoma procariotico. (A) Rappresentazione schematica del ribosoma 70S con indicati i siti A, P ed E per i tRNA, i centri attivi della peptidil transferasi nella subunità grande e di decodificazione nella subunità piccola, il canale dove scorre l'mRNA nella subunità piccola e il tunnel di uscita della proteina nascente attraverso la subunità grande. (B) Posizionamento nei suddetti siti dell'mRNA, dei tRNA e della catena proteica nascente.



subunità, cosicché ciascuno di essi è composto di due emisiti, uno nella subunità minore e uno nella subunità maggiore del ribosoma. La **figura 14.10B** mostra che i tRNA sono posizionati in modo che gli anticodoni, che si trovano all'estremo del braccio lungo della struttura a L dei tRNA (*vedi* oltre), si possano appaiare con i codoni dell'mRNA nella subunità minore del ribosoma in quello che viene chiamato **centro di decodificazione** (o "decifrazione"). Invece, le estremità del braccio corto della struttura a L dei tRNA che portano l'amminoacido e il peptide nascente, si trovano nella subunità maggiore e in particolare nel **centro della peptidil transferasi** (o "peptidil transferasico"), sito responsabile della formazione dei legami peptidici. Nella subunità 30S sono presenti due piccoli canali, uno di entrata e uno di uscita, in cui scorre l'mRNA durante il processo di traduzione. Il canale di entrata è di una larghezza tale per cui l'mRNA perde le sue strutture secondarie ed entra disteso nel centro di decodificazione, dove gli appaiamenti codone-anticodone nei siti A e P sono facilitati dal fatto che, tra i due codoni, l'mRNA forma un angolo. Nella subunità 50S si trova un altro canale, il tunnel per l'uscita della catena peptidica nascente che attraversa la subunità partendo dal centro peptidil transferasico. Questo tunnel è sufficientemente largo da permettere al peptide di assumere strutture ad α -elica, ma non strutture più ingombranti e complesse come foglietti β o interazioni terziarie, per le quali è necessario che la proteina neosintetizzata sia completata e rilasciata dal ribosoma.

Ribosoma come ribozima

Durante i primi decenni di studi sui ribosomi e sulla traduzione si considerava che le r-proteine fossero le vere responsabili della sintesi proteica e che gli rRNA costituissero essenzialmente delle impalcature necessarie a tenere insieme e posizionare correttamente le r-proteine. Successivamente però, con la scoperta della capacità dell'rRNA di catalizzare alcune reazioni chimiche e quindi dei primi ribozimi, si cominciò a far strada l'idea che gli rRNA non fossero solo degli spettatori

inerti nel processo di traduzione e che avessero loro, e non le r-proteine, il ruolo di catalizzatori della sintesi proteica. Questo concetto è stato poi confermato da due tipi di evidenze, una funzionale e una strutturale. L'evidenza funzionale è dovuta ai lavori iniziati negli anni novanta da Harry Noller (**fig. 14.8A**) e collaboratori, i quali dimostrarono sperimentalmente che l'rRNA 23S da solo, estratto dalla subunità 50S procariotica e privo di tutte le r-proteine, è in grado di catalizzare, seppure con una bassa efficienza, la formazione di legami peptidici. L'evidenza strutturale venne nell'anno 2000 con la definizione ad alta risoluzione della struttura del ribosoma. Infatti, come abbiamo già menzionato sopra, i due centri attivi del ribosoma, quello di decodificazione e quello peptidil transferasico, e anche l'interfaccia tra le due subunità, cioè i siti direttamente implicati nel processo di sintesi proteica, sono essenzialmente costituiti dagli rRNA, mentre vi mancano r-proteine, che sono invece distribuite soprattutto sulla parte esterna del ribosoma. Qual è allora il ruolo delle r-proteine? Interagendo con l'rRNA esse lo assistono nel ripiegarsi correttamente e nell'acquisire la corretta struttura tridimensionale definitiva. In questa maniera, pur senza prendere parte direttamente alla catalisi, contribuiscono sicuramente all'efficienza e all'accuratezza della traduzione. Come vedremo più avanti, la mancanza *in vivo* anche di una sola r-proteina è letale per la cellula. Si può ipotizzare, però, che in stadi molto precoci dell'evoluzione della vita sulla Terra ci possa essere stato uno stadio in cui una semplice sintesi di peptidi poteva essere catalizzata da "proto-ribosomi" costituiti di solo RNA che poi si sarebbero andati evolvendo in strutture sempre più complesse e accurate quali sono i ribosomi attuali.

14.2 RNA transfer (tRNA, RNA di trasferimento) e amminoacil-tRNA sintetasi

Verso la fine degli anni cinquanta, Francis Crick ipotizzò l'esistenza di "adattatori" molecolari che potessero riconoscere da una parte i codoni (triplette di nucleotidi) del messaggio genetico e dall'altra gli specifici amminoacidi da inserire nella sequenza delle proteine. Negli stessi anni era stata scoperta una classe di piccole molecole di RNA, circa il 10-15% dell'RNA totale, che si rivelarono ben presto gli adattatori ipotizzati e che sono ora chiamati **tRNA (RNA transfer, RNA di trasferimento)**.

In ogni cellula esiste almeno un tipo di tRNA per ciascuno dei 20 tipi di amminoacidi utilizzati nella sintesi proteica, ma spesso ve ne è più di uno. Convenzionalmente i tRNA per i diversi amminoacidi, ancorché non caricati, vengono indicati mettendo in apice il simbolo a tre lettere dell'amminoacido corrispondente: per esempio tRNA^{Leu} indica il tRNA specifico per la leucina, non caricato con l'amminoacido. Per indicare un tRNA carico, l'amminoacido caricato viene riportato come prefisso: per esempio Leu-tRNA^{Leu} indica un tRNA specifico per leucina caricato con l'amminoacido corrispondente. Come sarà discusso più avanti, è possibile che per errore, o sperimentalmente in laboratorio, un tRNA venga caricato con un amminoacido errato: potremmo avere così, per esempio, un Val-tRNA^{Ile} cioè un tRNA per isoleucina caricato con una valina.

I tRNA sono piccole molecole costituite da una sequenza di 74-95 nt. Anche i tRNA, come gli rRNA, sono caratterizzati dalla presenza di numerose basi insolite, che, come già detto, sono il prodotto di modificazioni enzimatiche post-trascrizionali. Alcune regioni della sequenza sono abbastanza conservate nei diversi tRNA, altre sono specifiche per ciascuno di essi; comunque, tutti i tRNA presentano una struttura complessiva comune. Come mostrato in **figura 14.11B**, i tRNA hanno una caratteristica struttura secondaria a trifoglio, un gambo e tre foglie (alcuni preferiscono chiamarlo quadrifoglio), risultante dall'appaiamento

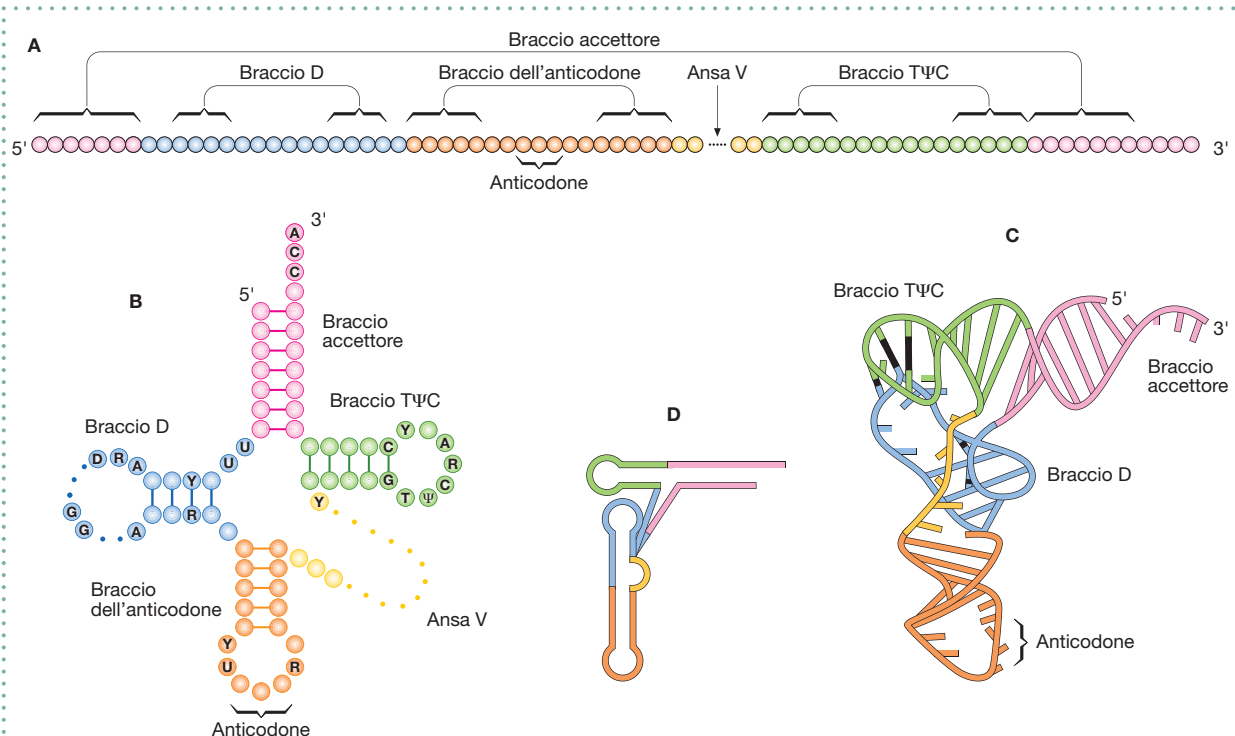


Figura 14.11 Struttura dei tRNA. (A) La struttura primaria è una sequenza di 74-95 nt che presenta delle regioni complementari qui indicate dalle coppie di graffe. (B) L'appaiamento delle regioni complementari fa assumere alla molecola la struttura secondaria a trifoglio. (C) La struttura secondaria a trifoglio si ripiega in una struttura terziaria compatta a forma di L, con il sito di legame per

l'amminoacido a un'estremità e l'anticodone all'altra estremità. (D) Questo schema aiuta a capire come si passa dalla struttura a quadrifoglio (B) alla struttura terziaria (C). Le varie regioni della molecola hanno lo stesso codice colori in (A), (B), (C) e (D).

di basi complementari che, come mostrato in **figura 14.11A**, si trovano in diverse regioni della sequenza nucleotidica. Nella struttura a trifoglio riconosciamo un braccio accettore formato dalle due estremità della molecola e tre bracci costituiti ciascuno da uno stelo a doppio filamento e da un'ansa di basi non appaiate, chiamati braccio TΨC, braccio D e braccio dell'anticodone. Talvolta può essere presente un braccio accessorio di lunghezza variabile.

- **Braccio accettore:** formato dall'appaiamento delle due estremità della molecola, è così chiamato perché su di esso si lega l'amminoacido. Come vedremo in dettaglio più avanti, l'amminoacido si lega all'estremità 3', che è sempre una A preceduta da due C. Come descritto nel capitolo 12 (*vedi* par. 12.2 e fig. 12.4), questa sequenza ...CCA-3' che protrude dallo stelo a doppio filamento di tutti i tRNA è, nei procarioti, codificata dal gene per il tRNA e quindi è presente nel tRNA precursore. Al contrario, i precursori dei tRNA eucariotici non contengono in genere la sequenza terminale CCA, che viene invece aggiunta all'estremità 3' da una tRNA nucleotidil transferasi.
- **Braccio TΨC:** così chiamato per la presenza, nella sua ansa a singolo filamento, di questa sequenza, dove T indica una timina (anomala per un RNA) e Ψ indica la base modificata pseudouridina.
- **Braccio D:** prende il nome dalla presenza nella sua ansa di un'altra base modificata, la diidrouridina.
- **Braccio dell'anticodone:** così chiamato per la presenza nella sua ansa della tripletta di nucleotidi che riconosce per appaiamento di basi il codone sull'mRNA e che si trova sempre fiancheggiata al lato 3' da una purina e al lato 5' da un uracile.

FINESTRA 14.1

QUANTI AMMINOACIDI SONO CODIFICATI DAL CODICE GENETICO?

Quanti sono i diversi tipi di amminoacidi? Il numero di 20 si riferisce agli amminoacidi normalmente inseriti nelle proteine durante la sintesi proteica. Alcuni di questi però possono poi subire delle modificazioni post-traduzionali, quali fosforilazione, formilazione, metilazione ecc., per cui nelle proteine mature possiamo trovare vari tipi di amminoacidi. Inoltre nella cellula esistono ancora altri amminoacidi non inseriti nelle proteine, magari perché intermedi nelle varie vie metaboliche ecc.

Anche il numero di 20 per gli amminoacidi codificati dal genoma è una regola generale che, come spesso succede, presenta rare ma molto interessanti eccezioni. Un caso ben studiato è quello dell'amminoacido modificato **selenocisteina** (Sec), che è una cisteina in cui l'atomo di zolfo è sostituito da un atomo di selenio. Le poche proteine, procariotiche ed eucariotiche, che contengono un residuo di selenocisteina vengono chiamate **selenoproteine**. Queste di solito catalizzano reazioni di ossido-riduzione e contengono un unico residuo di selenocisteina, che fa parte del sito attivo. L'esempio meglio studiato è quello di tre geni di *Escherichia coli* che codificano per isozimi della formato-deidrogenasi. La selenocisteina non è presente normalmente nella tabella del codice genetico, ma non è nemmeno il risultato di una modificazione post-traduzionale come sono le fosforilazioni, acetilazioni ecc. sopra menzionate. Infatti essa è codificata in un modo molto peculiare da un codone UGA, che normalmente è un codone di terminazione.

Perché alcuni codoni UGA codificano per la selenocisteina

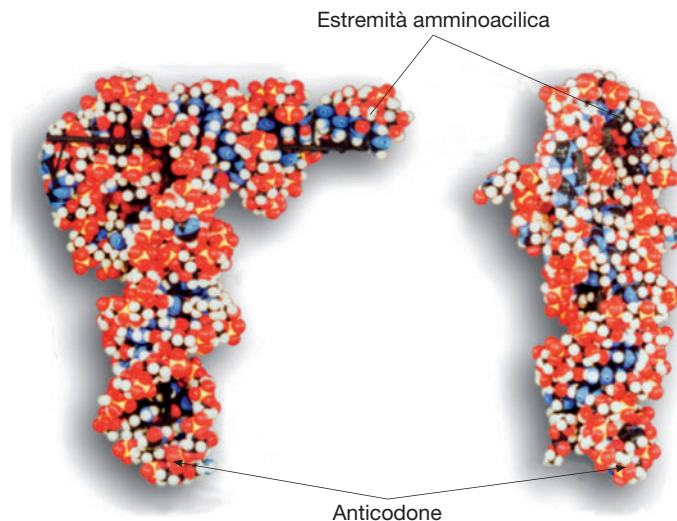
mentre la maggior parte di essi porta alla terminazione? La differenza sta nel "contesto", cioè un codone UGA viene letto come selenocisteina solo se è seguito da una particolare sequenza che forma una forcina (stelo e ansa) chiamata **elemento SECIS**. In *E. coli* questa struttura secondaria viene riconosciuta e legata dalla proteina SelB. Questa è costituita da una porzione che funziona da fattore di allungamento alternativo a EF-Tu, e, infatti, gran parte della sua sequenza è correlata a quella di EF-Tu. Inoltre SelB ha, rispetto a EF-Tu, un dominio extra che riconosce e lega l'elemento SECIS. Quindi selB riconosce specificamente gli mRNA per selenoproteine e media, solo per questi, l'ingresso del Seleno-Cys-tRNA nel sito A del ribosoma in corrispondenza di un codone UGA. Naturalmente non è sufficiente il gene per SelB perché il sistema funzioni. Infatti il sistema procariotico comprende anche il gene *selC*, che codifica per un particolare tRNA il cui anticodone 5'-UCA-3' si può appaiare con il codone 5'-UGA-3', e *selA* e *selD*, che codificano per due proteine implicate nella modificazione della cisteina, una volta caricata sul tRNA, in selenocisteina. Anche nei mammiferi alcuni codoni UGA sono decodificati come selenocisteine con un meccanismo sostanzialmente simile, anche se con alcune differenze.

Un ulteriore esempio è l'inserimento di un altro amminoacido, la **pirrolisina**, in una proteina di un batterio e di un archeobatterio, in corrispondenza di un altro codone di terminazione, UAG, probabilmente con un meccanismo analogo a quello qui descritto per la selenocisteina.

- **Braccio variabile:** o braccio extra, è situato tra il braccio TΨC e quello dell'anticodone ed è di lunghezza variabile da 3 a 21 basi nei diversi tRNA.

La struttura a trifoglio, anche se utile per mettere in evidenza come diverse regioni della molecola interagiscono per appaiamento di basi, non rappresenta la vera struttura dei tRNA. Infatti i quattro steli a doppio filamento, anche se brevi, tendono ad assumere la classica struttura a doppia elica e vanno nel loro insieme a formare una struttura tridimensionale complessa. Questa struttura è stata risolta mediante diffrazione dei raggi X. Perché questo fosse possibile è stato prima necessario purificare, dalla miscela di tutti i tRNA che si estraggono dalle cellule, un solo tipo di tRNA e ottenerne poi dei cristalli. I primi studi dettagliati sono stati fatti sul tRNA per la fenilalanina, tRNA^{Phe}, di lievito. I risultati ottenuti da questo e da altri studi hanno dimostrato che la struttura secondaria a trifoglio di tutti i tRNA è ripiegata su se stessa a formare una struttura terziaria (tridimensionale) cosiddetta a L. Questa è mostrata schematicamente in **figura 14.11C**, mentre la **figura 14.12** ne mostra un modello ad atomi pieni. Il confronto della **figura 14.11D** con la **14.11B** e la **14.11C** aiuta a capire come i quattro bracci della rappresentazione a trifoglio dei tRNA vanno a sistemarsi nella struttura tridimensionale. In ogni caso nella struttura a L troviamo l'anticodone all'estremità del braccio lungo della L e, a circa 70 Å di distanza, l'amminoacido caricato sull'estremità del braccio corto della L. I due bracci D e TΨC vanno a formare l'angolo della L, con le due anse non appaiate all'esterno dell'angolo. La

Figura 14.12 Struttura terziaria a L del tRNA. Modello ad atomi pieni del tRNA^{Phe} di lievito. Il tRNA è visto da due angolazioni ruotate di 90° una rispetto all'altra.



struttura a L è stabilizzata da alcuni legami idrogeno che si formano tra le basi non appaiate delle due anse ravvicinate, da alcune interazioni tra basi e scheletro zucchero-fosfato e infine da interazioni idrofobiche (impilamento) tra basi che si vengono a trovare sovrapposte.

In effetti i tRNA non possono da soli svolgere il ruolo di adattatori molecolari ipotizzato da Crick. Infatti, mentre essi sono in grado di riconoscere direttamente (seppure con l'aiuto del ribosoma) il codone mediante interazioni specifiche codone-anticodone, non hanno però alcuna diretta affinità specifica per gli amminoacidi che devono caricare. Il processo di caricamento degli amminoacidi sui corrispondenti tRNA viene catalizzato da una classe di enzimi, chiamati **amminoacil-tRNA sintetasi**, che così completano il ruolo di adattatori dei tRNA. In genere nella cellula ci sono 20 differenti tipi di amminoacil-tRNA sintetasi, una per ciascun tipo di amminoacido. Mentre ogni amminoacil-tRNA sintetasi riconosce specificamente un solo amminoacido, essa riconosce tutti i tRNA sui quali questo può essere caricato, cioè i cosiddetti **tRNA isoaccettori**. Il riconoscimento di diversi isoaccettori da parte di una sintetasi è garantito da contatti dell'enzima con un numero ridotto di basi (1-5) distribuite in varie posizioni della molecola di tRNA. In genere sono coinvolte almeno una base dell'anticodone, una delle ultime coppie di basi dello stelo accettore e la cosiddetta base discriminante adiacente al CCA terminale.

Il ruolo delle amminoacil-tRNA sintetasi è quello di catalizzare la formazione di un legame acilico ad alta energia tra il gruppo carbossilico dell'amminoacido e il gruppo idrossilico in posizione 2' o 3' della adenosina che costituisce l'estremità 3' del tRNA. Questa energia di legame è molto importante perché verrà successivamente utilizzata per la formazione del legame peptidico nel corso della sintesi proteica; essa è derivata dall'idrolisi di una molecola di ATP al momento dell'attivazione dell'amminoacido da parte della amminoacil-tRNA sintetasi stessa. Come illustrato in **figura 14.13**, il caricamento dell'amminoacido sul tRNA avviene sempre in due passaggi consecutivi, ambedue catalizzati dalla stessa amminoacil-tRNA sintetasi. Nel primo passaggio ($aa + ATP \rightarrow aaAMP + PP$) l'amminoacido reagisce con una molecola di ATP formando un ammino-aciladenilato (aa) e rilasciando pirofosfato; l'energia necessaria alla reazione è fornita dall'idrolisi del legame ad alta energia dell'ATP. Nel secondo passaggio ($aa\sim AMP + tRNA \rightarrow aa\sim tRNA + AMP$) l'amminoacido viene trasferito dall'AMP all'adenina 3' terminale del tRNA con rilascio di AMP. La **figura 14.14** mostra come, conseguen-

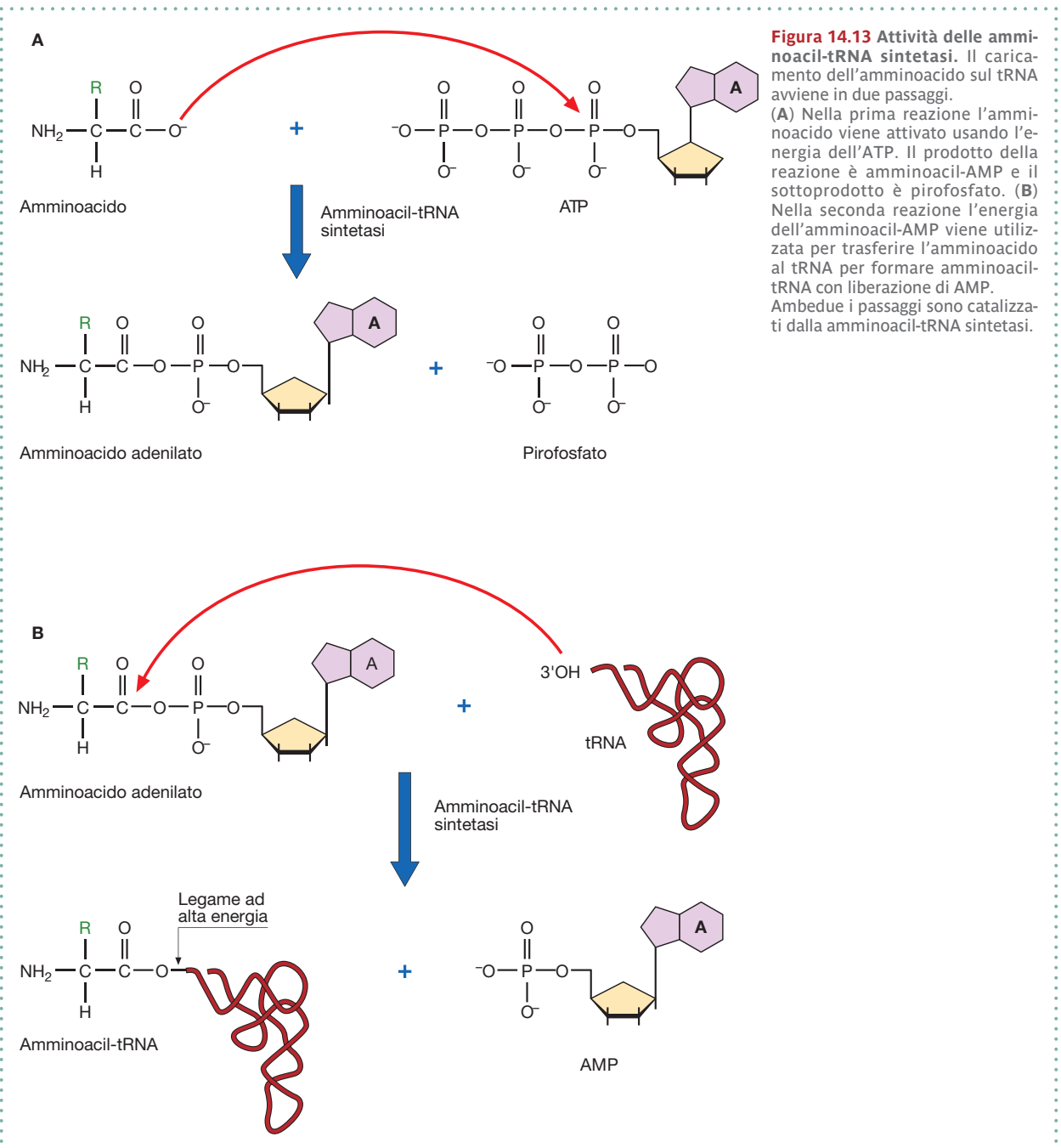
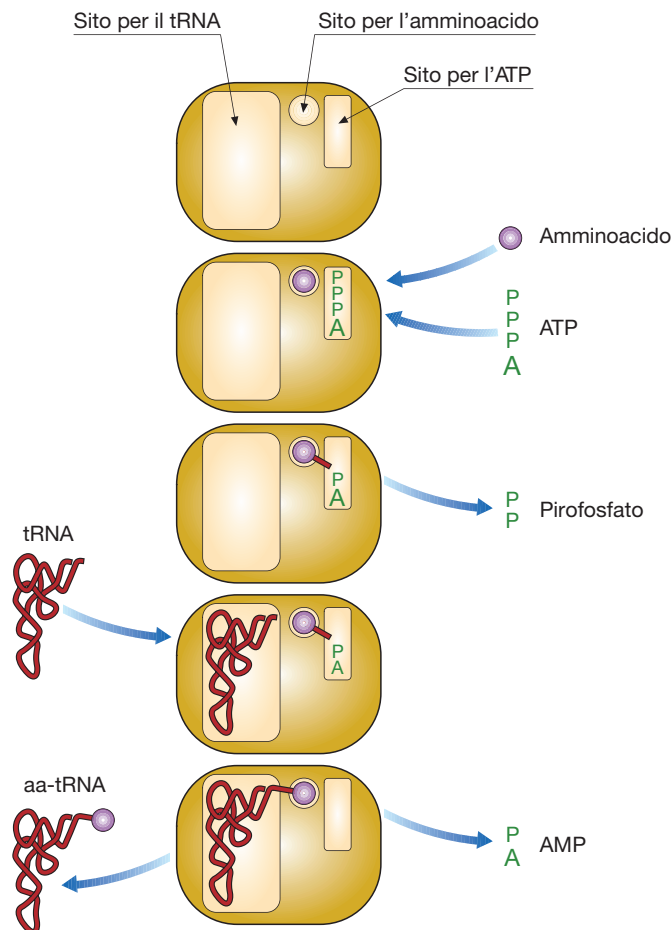


Figura 14.13 Attività delle amminoacil-tRNA sintetasi. Il caricamento dell'amminoacido sul tRNA avviene in due passaggi.

(A) Nella prima reazione l'amminoacido viene attivato usando l'energia dell'ATP. Il prodotto della reazione è amminoacil-AMP e il sottoprodotto è pirofosfato. (B) Nella seconda reazione l'energia dell'amminoacil-AMP viene utilizzata per trasferire l'amminoacido al tRNA per formare amminoacil-tRNA con liberazione di AMP. Ambedue i passaggi sono catalizzati dalla amminoacil-tRNA sintetasi.

temente, le amminoacil-tRNA sintetasi hanno sempre tre siti di legame: uno per l'amminoacido, uno per il tRNA corrispondente e uno per una molecola di ATP che è necessaria alla reazione. Le amminoacil-tRNA sintetasi costituiscono nella cellula un gruppo abbastanza eterogeneo di enzimi, anche se svolgono la stessa funzione. Alcune sono monomeriche, altre dimeriche o tetrameriche, e il peso molecolare del monomero può variare da 40 a 110 kDa. Sulla base della struttura del sito attivo, le sintetasi possono essere distinte in due gruppi non correlati che potrebbero avere avuto origini e storie evolutive indipendenti. Le **sintetasi di classe I**, generalmente monomeriche, legano inizialmente l'amminoacido al 2'OH del tRNA, mentre le **sintetasi di classe II**, generalmente dimeriche o tetrameriche, legano l'amminoacido al 3'OH del tRNA. Comunque, dopo che le

Figura 14.14 Rappresentazione schematica di una amminoacil-tRNA sintetasi con i tre siti di legame per l'ATP, per l'amminoacido e per il corrispondente tRNA. L'enzima catalizza prima l'attivazione dell'amminoacido con ATP con formazione di amminoacil-AMP. Successivamente la stessa sintetasi catalizza il trasferimento dell'amminoacido attivato al tRNA.



sintetasi, sia di classe I che di classe II, si sono staccate, l'amminoacido raggiunge presto un equilibrio tra legame al 2'OH e 3'OH.

Questa doppia reazione è molto complessa anche perché è essenziale, per l'accuratezza della sintesi delle proteine, che ogni tRNA sia caricato correttamente con l'amminoacido corrispondente. Quindi ogni sintetasi deve riconoscere con molta specificità sia l'amminoacido che i suoi corrispondenti tRNA isoaccettori, cosa non affatto banale. Infatti, alcuni amminoacidi sono strutturalmente molto simili tra loro e quindi non facilmente discernibili. D'altra parte anche i tRNA, pur essendo diversi per sequenza, hanno tutti una struttura tridimensionale molto simile.

A tal fine entrano in gioco dei meccanismi di "correzione di bozze" (*proofreading*) per cui i due processi, riconoscimento e caricamento del corretto amminoacido e riconoscimento del corrispondente tRNA da parte della sintetasi, vengono ambedue svolti e rivisti in stadi successivi in modo da ridurre la probabilità di errore. In altre parole, le sintetasi controllano il risultato della reazione a livello di alcuni o di tutti i suoi stadi e ne annullano l'esito quando sia stato usato per la reazione il tRNA o l'amminoacido sbagliato. Dei due processi, riconoscimento dell'amminoacido e riconoscimento del tRNA, il primo è il più difficile e più soggetto a errori. Infatti, essendo gli amminoacidi molto più piccoli dei tRNA, hanno meno possibilità di interazioni specifiche con le corrispondenti sintetasi. Inoltre, come abbiamo già detto, alcuni amminoacidi sono estremamente simili tra loro. In genere le sintetasi escludono dal sito di attivazione alcuni amminoacidi perché troppo grandi o con forme diverse da quella corretta, ma possono erroneamente accettare un amminoacido errato strutturalmente simile a quello

corretto. In questo caso la frequenza di errore viene però ridotta dal passaggio dell'aa-AMP, prima del trasferimento sul tRNA, in un "sito di correzione" presente nell'enzima stesso. Riportiamo qui l'esempio meglio studiato, quello della amminoacil-sintetasi per l'isoleucina di *E. coli*. Questa sintetasi esclude dal sito di attivazione la fenilalanina perché troppo grande e la leucina perché di forma non compatibile, ma può accettare la valina e attivarla, cioè legarla all'AMP. L'amminoacido attivato però, prima di essere trasferito sul tRNA, transita nel sito di correzione dell'enzima stesso che, se l'amminoacido è quello corretto (in questo caso isoleucina), lascia proseguire la reazione, mentre se coinvolge un amminoacido errato (come la valina) porta a idrolisi del Val-AMP con liberazione di valina e AMP. In alcuni casi una sintetasi non solo lega e attiva un amminoacido errato ma lo trasferisce anche sul tRNA, che, essendo quello riconosciuto dall'enzima, non sarà quello corretto per l'amminoacido. Una volta formato, questo amminoacil-tRNA viene riconosciuto come aberrante dalla sintetasi che lo idrolizza.

Qual è il senso di tutto questo? In genere, reazioni chimiche complesse in cui sia richiesta un'alta specificità di riconoscimento molecolare possono essere soggette a tassi di errore non trascurabili. Mentre in alcuni casi questo "rumore di fondo" può non essere particolarmente dannoso, esso può essere intollerabile in casi specifici, come per esempio la replicazione del DNA e la sintesi proteica. Naturalmente dovrebbe essere possibile l'evoluzione di sistemi di sintesi così sofisticati da avere un tasso di errore estremamente basso, ma sicuramente richiederebbero apparati sintetici molto più complessi e un dispendio di energia molto elevato. È molto più economico verificare la reazione in due o più passaggi, anche se ciascuno con un tasso di errore relativamente alto. La seguente analogia, anche se un po' banale, spiega bene il concetto: per una casa editrice è più economico, in termini di tempo e di soldi, comporre rapidamente una prima "bozza" di un testo e verificarlo poi mediante una "correzione di bozze", piuttosto che non comporlo subito in forma assolutamente corretta. Tornando al caso sopra descritto del tRNA^{Ile}, se l'errore della prima fase di reazione, cioè l'attivazione dell'amminoacido, è di 1/225 e l'errore della reazione di correzione è di 1/270, si ha che l'errore complessivo di formazione di un aberrante Val-tRNA^{Ile} è di $1/225 \times 1/270 = 1/60\,000$.

14.3 RNA messaggero (mRNA)

Nel loro fondamentale lavoro del 1961 sulla regolazione dell'operone *lac*, François Jacob e Jacques Monod escludevano, sulla base di dati sperimentali, che l'rRNA con la sua grande stabilità metabolica potesse svolgere il ruolo di portatore dell'informazione per la sintesi delle proteine e ipotizzarono quindi l'esistenza di una distinta classe di molecole di RNA metabolicamente instabili, che chiamarono RNA messaggero, la cui vita media molto breve permettesse alla cellula di rispondere rapidamente a cambiamenti ambientali adattando la produzione di proteine alle esigenze del momento.

Così, delle tre classi di RNA coinvolti nella sintesi proteica l'mRNA è quello che è stato scoperto per ultimo, anche perché più difficile da isolare e analizzare rispetto agli altri due componenti, rRNA e tRNA. Infatti, l'mRNA rappresenta quantitativamente una bassa percentuale degli RNA cellulari (2-5% rispetto all'80% di RNA ribosomale e al 15% del tRNA). Inoltre, poiché una cellula produce migliaia di mRNA diversi per codificare le migliaia di differenti proteine che la caratterizzano, la popolazione di molecole di mRNA è estremamente eterogenea per grandezza e sequenza nucleotidica. Come previsto da Jacob e Monod, gli mRNA hanno, a confronto con gli rRNA e tRNA, una vita media molto breve. Nei procarioti molti mRNA hanno una vita media di solo 2 minuti, cioè un decimo del tempo di duplicazione cellulare. Come mostrato in **figura 14.15**, questa alta instabilità, insieme al fatto che nei procarioti la traduzione di

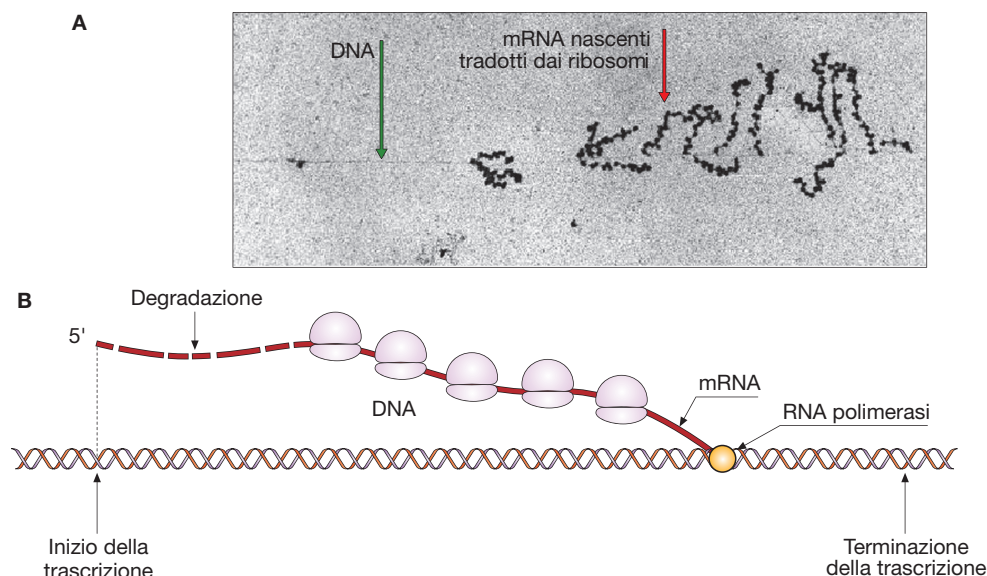


Figura 14.15 Relazione tra processi di trascrizione, traduzione e degradazione degli mRNA nei procarioti. (A) Unità di trascrizione visualizzata con il microscopio elettronico (foto: Oscar L. Mil-

ler). (B) Rappresentazione schematica di una unità di trascrizione in cui si vede la contemporanea trascrizione, traduzione e degradazione di un mRNA.

un mRNA è contestuale alla sua trascrizione, porta a una curiosa conseguenza: di alcuni tipi di mRNA, soprattutto quelli lunghi, non vi sono mai molecole intere. Infatti, mano a mano che sono trascritti dal 5' verso il 3' vengono tradotti dai ribosomi che avanzano nella stessa direzione ma, in alcuni casi, possono anche iniziare a essere degradati al 5' quando ancora non è terminata la trascrizione all'estremità 3'. Nelle cellule eucariotiche la situazione è diversa. Infatti, trascrizione e traduzione avvengono in compartimenti separati, nucleo e citoplasma. Gli mRNA, maturati e trascritti nel nucleo e poi esportati nel citoplasma per la traduzione, hanno una vita media più lunga rispetto agli mRNA procariotici, ma comunque molto inferiore a quella degli altri RNA cellulari. Inoltre, questa vita media è molto variabile per i diversi mRNA e in diverse situazioni e, come vedremo più avanti, fornisce opportunità per importanti processi regolativi.

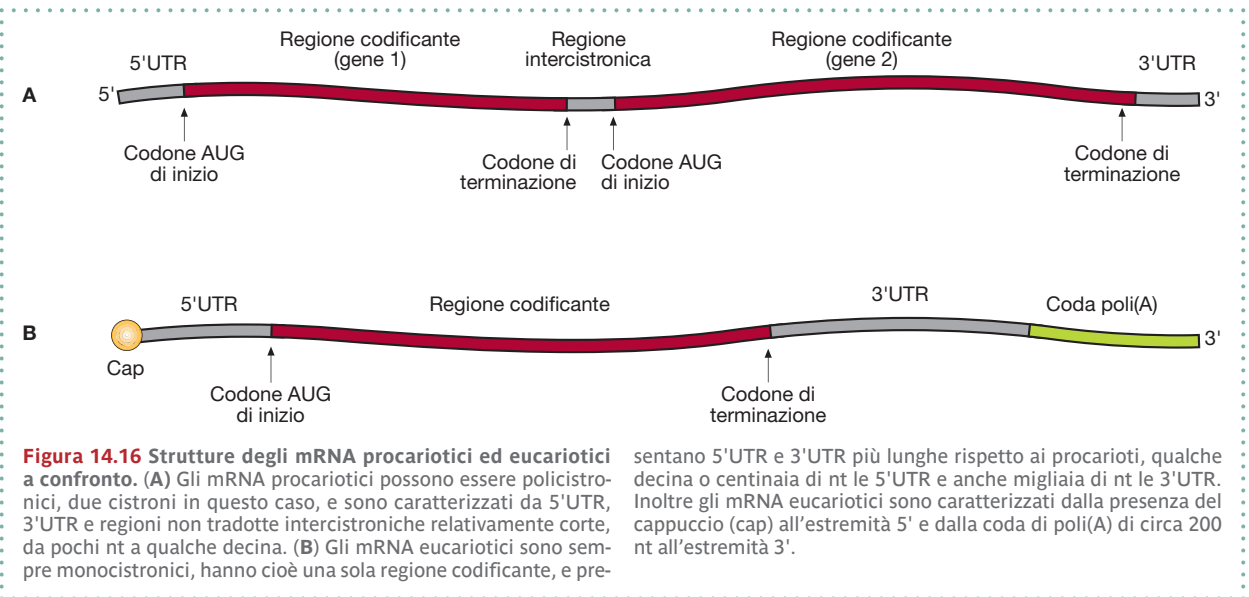
La **figura 14.16** mette a confronto le strutture di un tipico mRNA procariotico e uno eucariotico. Si chiama **regione codificante** di un mRNA quella che

FINESTRA 14.2

LA CODA DI POLI(A) COME MANIGLIA PER PURIFICARE L'mRNA

La presenza all'estremità 3' degli mRNA eucariotici di una coda di poli(A), lunga circa 200 residui di adenina, viene sfruttata in laboratorio per purificare questa classe di RNA dall'RNA totale preparato da cellule o tessuti. Infatti si possono preparare piccole colonne cromatografiche in cui la matrice della colonna (ad es. cellulosa o sefariosio) contenga legati covalentemente degli oligo(dT), oligonucleotidi costituiti da sequenze di dT (deossitimidina). Al passaggio attraverso la colonna di una soluzione contenente RNA totale, solo le molecole di mRNA si legheranno con la loro coda di poli(A) alle sequenze di oligo(dT), mentre la maggior parte degli altri RNA (rR-

NA, tRNA ecc.) non sarà trattenuta e potrà essere lavata via. Dopo gli opportuni lavaggi, sarà sufficiente innalzare un po' la temperatura e/o abbassare la forza ionica della soluzione per eluire l'mRNA precedentemente legato. Questa frazione purificata di RNA viene chiamata **RNA poli(A)⁺**, termine che viene spesso usato come sinonimo di mRNA. In realtà l'equivalenza poli(A)⁺ = mRNA non è in assoluto corretta poiché esistono alcuni rarissimi mRNA che non hanno coda di poli(A) e, viceversa, possono esistere alcuni RNA non mRNA che hanno nella loro sequenza dei tratti di adenine abbastanza lunghi da essere trattenuti dalla colonna cromatografica.



sentano 5'UTR e 3'UTR più lunghe rispetto ai procarioti, qualche decina o centinaia di nt le 5'UTR e anche migliaia di nt le 3'UTR. Inoltre gli mRNA eucariotici sono caratterizzati dalla presenza del cappuccio (cap) all'estremità 5' e dalla coda di poli(A) di circa 200 nt all'estremità 3'.

codifica per la proteina ed è quindi caratterizzata da una ORF (*Open Reading Frame*, fase di lettura aperta). Questa comincia con un codone di inizio seguito da una serie di codoni “senso” che codificano per gli amminoacidi e termina con un codone di stop o terminazione. Come mostrato nella figura, gli mRNA procariotici possono essere policistronici, cioè possono essere il prodotto della trascrizione di più geni adiacenti nel genoma (operoni) e codificare quindi per più proteine. Gli mRNA eucariotici sono invece quasi sempre monocistronici, cioè ogni mRNA codifica per una sola proteina e contiene quindi una sola ORF. Inoltre, negli mRNA, sia procariotici che eucariotici, sono quasi sempre presenti anche regioni non codificanti chiamate **UTR** (*UnTranslated Region*, regione non tradotta). All'estremità 5' abbiamo una 5'UTR, chiamata anche **sequenza leader**, mentre all'estremità 3' abbiamo una 3'UTR, chiamata anche **sequenza trailer**. Le 5'UTR e 3'UTR sono in genere molto corte nei procarioti, da pochi a qualche decina di nucleotidi, e più lunghe negli eucarioti, dove si possono avere 3'UTR anche di migliaia di nucleotidi. Nel caso degli mRNA policistronici dei procarioti sono presenti anche una o più regioni non codificanti intergeniche (usualmente piuttosto corte, al massimo qualche decina di nucleotidi) che separano le regioni che codificano per le varie proteine.

Quasi tutti gli mRNA eucariotici, ma non quelli procariotici, sono caratterizzati invece da particolari modificazioni alle due estremità 5' e 3': il cappuccio (o “cap” o “5'cap”) all'estremità 5' e la coda di poli(A) all'estremità 3'. Di queste strutture e di come vengono aggiunte agli mRNA eucariotici abbiamo già parlato nei paragrafi 12.5 e 12.6.

Lecture di approfondimento

- Korostelev, A. e Noller, H.F. (2007), «The ribosome in focus: new structures bring new insights», *Trends in Biochemical Sciences*, **32**, pp. 434-441.
- Ling, J., Reynolds, N. e Ibba, M. (2009), «Aminoacyl-tRNA synthesis and translational quality control», *Annual Review of Microbiology*, **63**, pp. 61-78.
- Moore, P.B. e Steitz, T.A. (2005), «The ribosome revealed», *Trends in Biochemical Sciences*, **30**, pp. 281-283.
- Noller, H.F., Hoang, L. e Fredrick, K. (2005), «The 30S ribosomal P site: a function of 16S rRNA», *FEBS Letters*, **579**, pp. 855-858.

15

Meccanismo della sintesi proteica

15.1 Inizio nei procarioti

15.2 Inizio negli eucarioti

15.3 Allungamento

15.4 Terminazione

15.5 Bilancio energetico

15.6 Velocità e accuratezza della sintesi proteica

15.7 Traduzione a livello strutturale

In questo capitolo vedremo come l'interazione tra i vari componenti dell'apparato di traduzione descritti precedentemente risulta nella "traduzione" dell'informazione genetica portata dalla sequenza di basi dell'RNA messaggero nella sequenza di amminoacidi del prodotto proteico. Questo processo di traduzione è alquanto complesso ed è molto conservato in tutti gli organismi. Prenderemo in considerazione prima l'inizio, poi l'allungamento e infine la terminazione della traduzione, confrontando sempre i meccanismi operanti nei procarioti e negli eucarioti. Manterremo la descrizione del processo a un livello abbastanza generico limitandoci a descrivere le interazioni tra i vari componenti. Alla fine del capitolo mostreremo, con alcuni esempi, come lo studio delle interazioni tra i componenti del processo di traduzione stia entrando in una fase di studio strutturale a livello più fine.

Quello che sappiamo sul complesso meccanismo della sintesi proteica è dovuto al lavoro di numerosi gruppi di ricerca. La **figura 15.1** mostra solo alcuni dei maggiori protagonisti di queste ricerche.

L'evento fondamentale della sintesi proteica è la formazione di un legame peptidico tra il gruppo amminico di un nuovo amminoacido e il gruppo carbossilico della catena peptidica nascente. Al momento della formazione del legame peptidico la catena peptidica nascente e il nuovo amminoacido si trovano ambedue legati a tRNA posizionati rispettivamente nel sito P e nel sito A del ribosoma (**fig. 15.2**). Il legame peptidico si forma senza previa rottura del legame tra l'amminoacido e il tRNA. Infatti nel centro peptidil transferasico del ribosoma le estremità dei due tRNA si trovano ravvicinate tra loro in modo che il gruppo amminico del nuovo amminoacido possa attaccare il gruppo carbossilico dell'ultimo amminoacido del peptide nascente che è legato al 3' dell'altro tRNA. Quindi, il peptide nascente che si trovava legato a un tRNA viene trasferito su un altro tRNA con l'interposizione del nuovo amminoacido. Per questa ragione la reazione fondamentale della sintesi proteica viene chiamata **reazione della peptidil**



Figura 15.1 Alcuni dei principali ricercatori che hanno contribuito a chiarire i meccanismi della sintesi proteica. (A) Masayasu Nomura (1927-2011), University of California-Irvine, USA; (B) Nahum Sonenberg, McGill University, Montreal, Canada; (C) John W.B. Hershey, University of California-Davis, USA.

transferasi o peptidil transferasica. Dopo il trasferimento, il peptide nascente si trova legato al tRNA nel sito A, ma il successivo processo di traslocazione ripristina la situazione iniziale con il peptidil-tRNA nel sito P e il sito A libero per accogliere un nuovo aa-tRNA. Questa reazione si ripete ciclicamente durante la fase di allungamento della traduzione, mentre i processi che avvengono durante le fasi di inizio e di terminazione sono chiaramente diversi.

Come illustrato in **figura 15.3**, la subunità minore è la prima che interagisce con l'mRNA per formare il complesso di inizio; solo con l'aggiunta della subunità maggiore la sintesi della proteina può effettivamente iniziare. Il ribosoma intero scorre poi lungo la regione codificante dell'mRNA decodificando il messaggio e sintetizzando la catena peptidica, fino a quando incontra un codone di terminazione. Qui la proteina neosintetizzata viene rilasciata e il ribosoma si stacca dall'mRNA ritornando nel pool cellulare di ribosomi liberi. Questi devono essere dissociati nelle due subunità, che possono così essere riutilizzate per successivi cicli di traduzione.

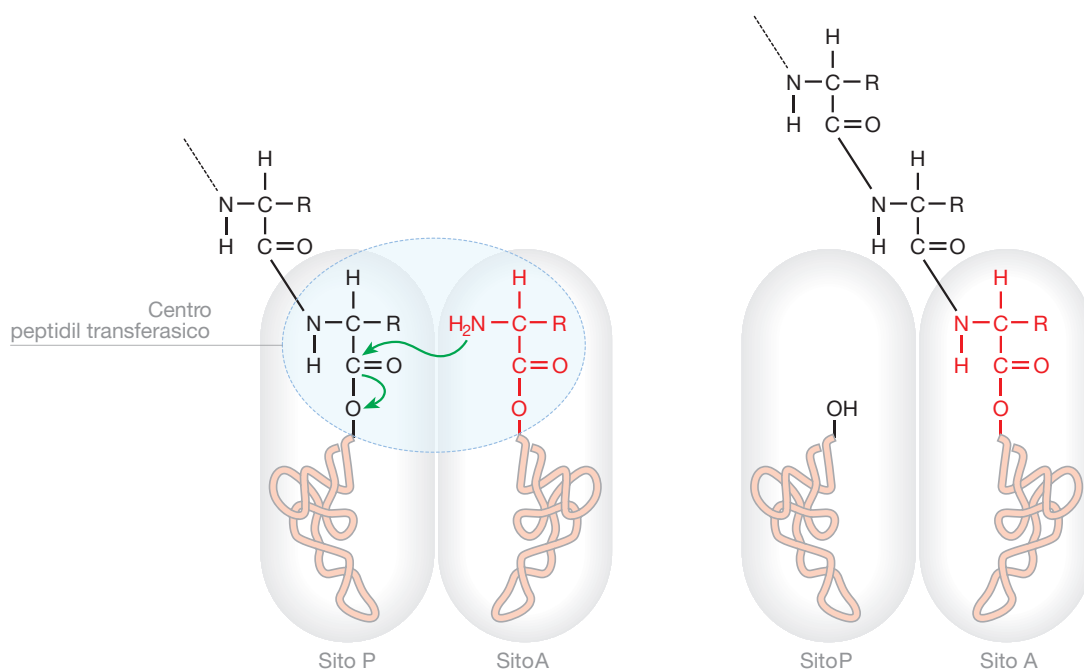
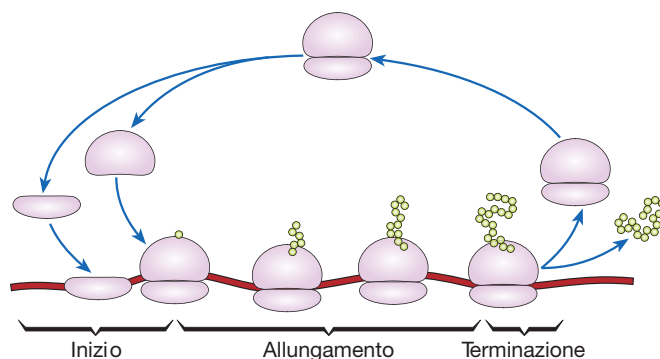


Figura 15.2 Reazione della peptidil transferasi. Nel centro peptidil transferasico del ribosoma, il gruppo amminico del nuovo aminoacido (rosso), legato al tRNA posizionato nel sito A, attacca

il gruppo carbossilico dell'ultimo aminoacido della catena peptidica nascente (nero) che si trova legato al tRNA posizionato nel sito P.

Figura 15.3 Le subunità ribosomali vengono riciclate e riutilizzate per numerosi cicli di traduzione. Nella fase di inizio l'mRNA lega la subunità minore, a cui poi si unisce anche la subunità maggiore. Il ribosoma scorre sull'mRNA decodificando la sequenza nucleotidica e sintetizzando la proteina. Giunto al codone di terminazione il ribosoma rilascia la proteina completata, si dissocia dall'mRNA e rientra nel pool di ribosomi liberi. Questi devono essere dissociati nelle due subunità per potere essere riutilizzati per altri cicli di traduzione.



Descriveremo ora in dettaglio gli eventi che caratterizzano l'inizio, l'allungamento e la terminazione della traduzione, nei procarioti e negli eucarioti. La maggior parte delle informazioni disponibili si riferiscono a pochi organismi modello: *Escherichia coli* per i procarioti, lievito e mammiferi per gli eucarioti. Come vedremo, la maggior differenza tra procarioti ed eucarioti si riscontra nella fase di inizio della traduzione, forse perché è la fase dove maggiormente agiscono i meccanismi di regolazione della sintesi proteica, meccanismi di cui gli eucarioti fanno massimo uso. Va precisato anche che per "sintesi proteica eucariotica" intendiamo quella sostenuta dal principale apparato di traduzione citoplasmatico. Non va però trascurato che le cellule eucariotiche contengono altri e importanti apparati di traduzione localizzati negli organelli, quali i mitocondri di tutti gli eucarioti e i cloroplasti delle cellule vegetali. Non parleremo qui di questi apparati di traduzione particolari; sottolineiamo solo che per molti aspetti essi si avvicinano di più alla traduzione procariotica che a quella eucariotica principale.

15.1 Inizio nei procarioti

tRNA di inizio

Sia nei procarioti che negli eucarioti la sintesi di tutte le proteine inizia all'estremità ammino-terminale con l'amminoacido metionina. Questa è codificata quasi sempre dal codone AUG, anche se, in alcuni rari casi, vengono usati anche i codoni GUG e UUG. Nei procarioti questa metionina di inizio è modificata chimicamente per aggiunta di un gruppo formilico che ne blocca il gruppo amminico, come mostrato in **figura 15.4**. La formilazione della metionina avviene ad opera dell'enzima metionil-tRNA transformilasi dopo essere stata caricata sul tRNA. Poiché durante il processo di sintesi proteica il gruppo amminico degli amminoacidi è essenziale per la formazione del legame peptidico con l'amminoacido precedente, la presenza del gruppo formilico impedisce alla metionina modificata di essere utilizzata nella fase di allungamento, ma non impedisce che venga utilizzata come primo amminoacido della catena. D'altra parte, residui di metionina sono presenti anche in posizioni interne alle sequenze proteiche. Infatti, nella cellula ci sono due tipi di tRNA per la metionina: il tRNA di inizio $\text{tRNA}_i^{\text{Met}}$ per la metionina che poi verrà formilata e il tRNA^{Met} per le metionine che saranno inserite internamente alle catene peptidiche. Quanto detto farebbe supporre che tutte le proteine cellulari abbiano all'estremità amminica una formil-metionina, ma sappiamo che ciò non è sempre vero. O meglio, ciò è vero solo al momento dell'inizio della sintesi della proteina, ma durante la fase di allungamento o dopo il completamento della proteina specifici eventi di maturazione post-traduzionale possono risultare, per alcune proteine, nella rimozione del gruppo formilico o

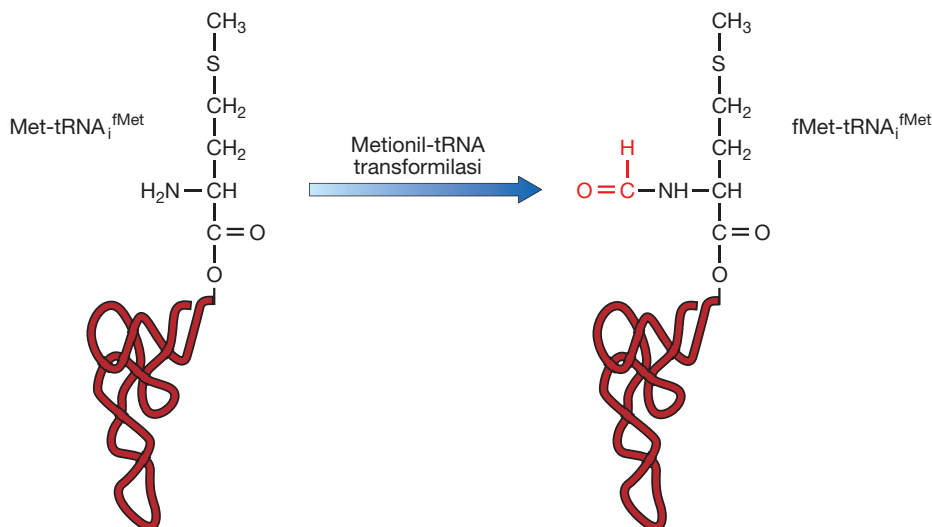


Figura 15.4 fMet-tRNA_i^{fMet} (N-formil-metionil-tRNA_i^{fMet}) si genera per formilazione di Met-tRNA_i^{fMet}. La reazione è catalizzata dalla metionil-tRNA transformilasi con formil-tetraidrofolato come cofattore.

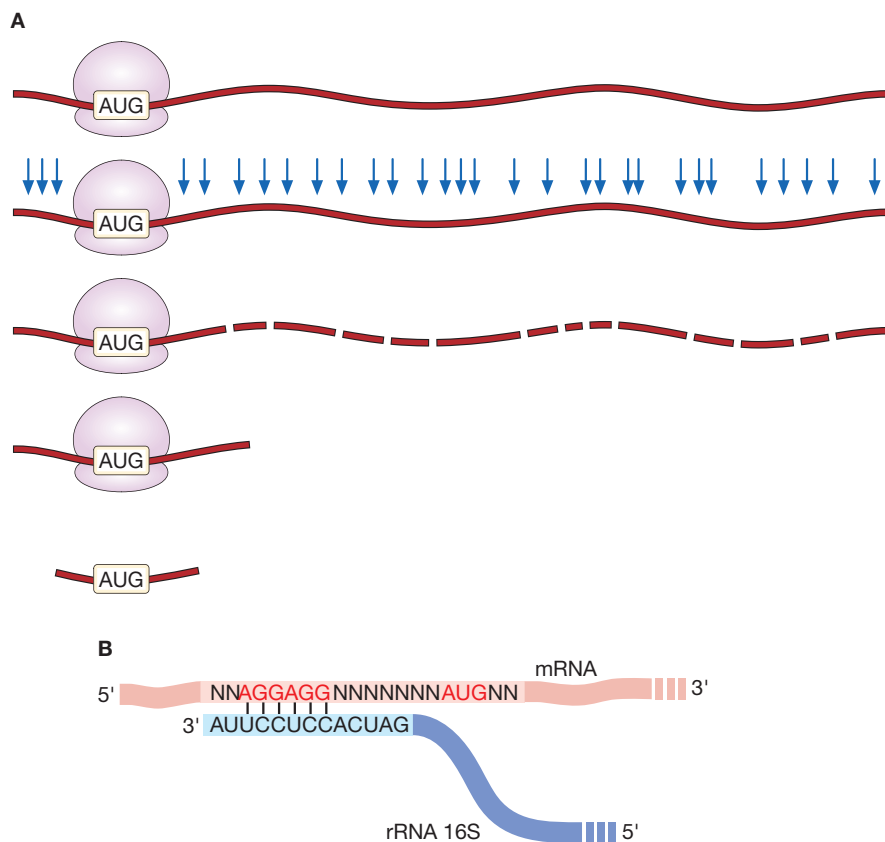
anche di uno o più amminoacidi. Un'altra particolarità del tRNA di inizio è che mentre tutti gli altri amminoacil-tRNA entrano nel sito A del ribosoma per passare poi, una volta convertiti in peptidil-tRNA, nel sito P, il tRNA di inizio fMet-tRNA_i^{fMet} entra direttamente nel sito P, anzi nell'emisito P, cioè nella regione della subunità 30S che andrà poi a formare il sito P nel ribosoma intero.

Riconoscimento del sito di inizio sull'mRNA

L'inizio della traduzione coinvolge sempre l'interazione preliminare tra la subunità minore del ribosoma, fMet-tRNA_i^{fMet}, e l'mRNA. Questi componenti vanno a costituire il complesso d'inizio 30S; solo successivamente si ha il reclutamento della subunità maggiore del ribosoma e la formazione del complesso di inizio 70S che poi procede nell'allungamento. È necessario che il complesso di inizio si assembli sull'mRNA in corrispondenza del codone di inizio, che è in genere un codone AUG. Possono ovviamente esserci altri codoni AUG nella fase di lettura aperta (ORF) che porta l'informazione dell'mRNA che codificano per metionine interne alla proteina sintetizzata; e ce ne possono essere anche nelle altre due fasi di lettura, quelle che non vengono lette come codoni durante la sintesi proteica. Come viene selezionato un vero codone AUG di inizio in corrispondenza del quale si deve assemblare il complesso di inizio? È stato possibile isolare e analizzare la sequenza della regione di mRNA su cui si assembla il complesso di inizio e che corrisponde al sito di legame al ribosoma. Infatti, come mostrato in **figura 15.5A**, facendo interagire *in vitro* ribosomi, mRNA, tRNA di inizio e GTP è stato possibile ottenere complessi di inizio che però restano bloccati non potendo procedere nell'allungamento. Questi complessi vengono digeriti con endonucleasi che degradano tutto l'mRNA tranne le regioni protette dai ribosomi, permettendo così di determinare la loro sequenza nucleotidica. Il confronto delle sequenze dei siti di legame al ribosoma di diversi mRNA ha messo in evidenza che, pochi nucleotidi a monte dell'AUG di inizio, è presente una sequenza molto conservata. In particolare, come mostrato in **figura 15.5B**, la sequenza AGGAGG precede, 7 nt a monte, il codone di inizio AUG. Questa sequenza rappresenta il sito di legame dei ribosomi all'mRNA e viene comunemente chiamata **sequenza di Shine-Dalgarno**, dai nomi dei due

Figura 15.5 Identificazione del “sito di legame per il ribosoma” (sequenza di Shine-Dalgarno) dell'mRNA procariotico.

(A) Una preparazione di mRNA veniva fatta interagire con ribosomi in condizioni in cui questi restavano bloccati sul sito di inizio. Successivamente, il trattamento con un'endonucleasi risultava nella degradazione di tutto l'mRNA tranne le brevi regioni protette dai ribosomi. Queste venivano recuperate e ne veniva analizzata la sequenza nucleotidica. (B) Il confronto delle sequenze nucleotidiche di molti di questi siti di legame per il ribosoma ha messo in evidenza la presenza, 7 nucleotidi a monte del codone di inizio AUG, di una sequenza conservata il cui consenso è AGGAGG. Tale sequenza, chiamata sequenza di Shine-Dalgarno, è complementare all'estremità 3' dell'rRNA 16S, che fa parte della subunità piccola del ribosoma.



ricercatori australiani John Shine e Lynn Dalgarno che l'hanno identificata. Questi ricercatori hanno anche dimostrato che essa è complementare all'estremità 3' dell'rRNA 16S, che è un costituente della subunità 30S del ribosoma (**fig. 15.5B**). In realtà, la sequenza di Shine-Dalgarno (AGGAGG) rappresenta un “consenso”, cioè una sequenza media rispetto alla quale ciascun mRNA può in parte deviare. Infatti, l'appaiamento anche di solo 4 o 5 basi con l'rRNA è sufficiente per l'interazione specifica tra mRNA e subunità ribosomale 30S e, quindi, per l'assemblaggio del complesso di inizio nella posizione corretta. Naturalmente solo i veri codoni AUG di inizio sono preceduti dalla sequenza Shine-Dalgarno. Gli altri AUG, per metionine interne o che si trovano nelle due fasi di lettura chiuse, non presentano Shine-Dalgarno e non sono utilizzati per iniziare la sintesi proteica. È da notare che un sito di inizio di questo tipo può trovarsi anche in posizioni interne lungo l'mRNA, cosicché una molecola di mRNA può avere anche più sequenze codificanti, ciascuna fornita di un suo sito di legame al ribosoma (Shine-Dalgarno+AUG) e di un suo codone di terminazione, il che avviene negli mRNA policistronici.

Fattori di inizio nei procarioti

Per il processo di inizio sono necessari, oltre all'mRNA, ai ribosomi e al tRNA di inizio caricato con la metionina, anche vari fattori proteici, che nei procarioti sono solo tre: **IF1**, **IF2** e **IF3** (IF, *Initiation Factor*). Tutti e tre i fattori d'inizio interagiscono con la subunità minore del ribosoma e sono necessari per la formazione del complesso di inizio 30S, che include, oltre a una subunità ribosomale 30S, anche un fMet-tRNA^{fMet} e un mRNA.

- IF3 è coinvolto nella dissociazione dei ribosomi interi nelle due subunità e successivamente, legato alla subunità 30S, è essenziale perché questa si leghi in modo specifico al sito di inizio sull'mRNA.
- IF1 contribuisce con IF3 alla dissociazione del ribosoma 70S e si lega alla subunità 30S nella regione che andrà a formare il sito A del ribosoma.
- IF2 è una GTPasi (cioè lega e idrolizza GTP per svolgere il suo ruolo) che interagisce con la subunità 30S, promuove l'associazione con $\text{fMet-tRNA}_i^{\text{fMet}}$ all'emisito P e impedisce l'interazione con altri tRNA carichi.

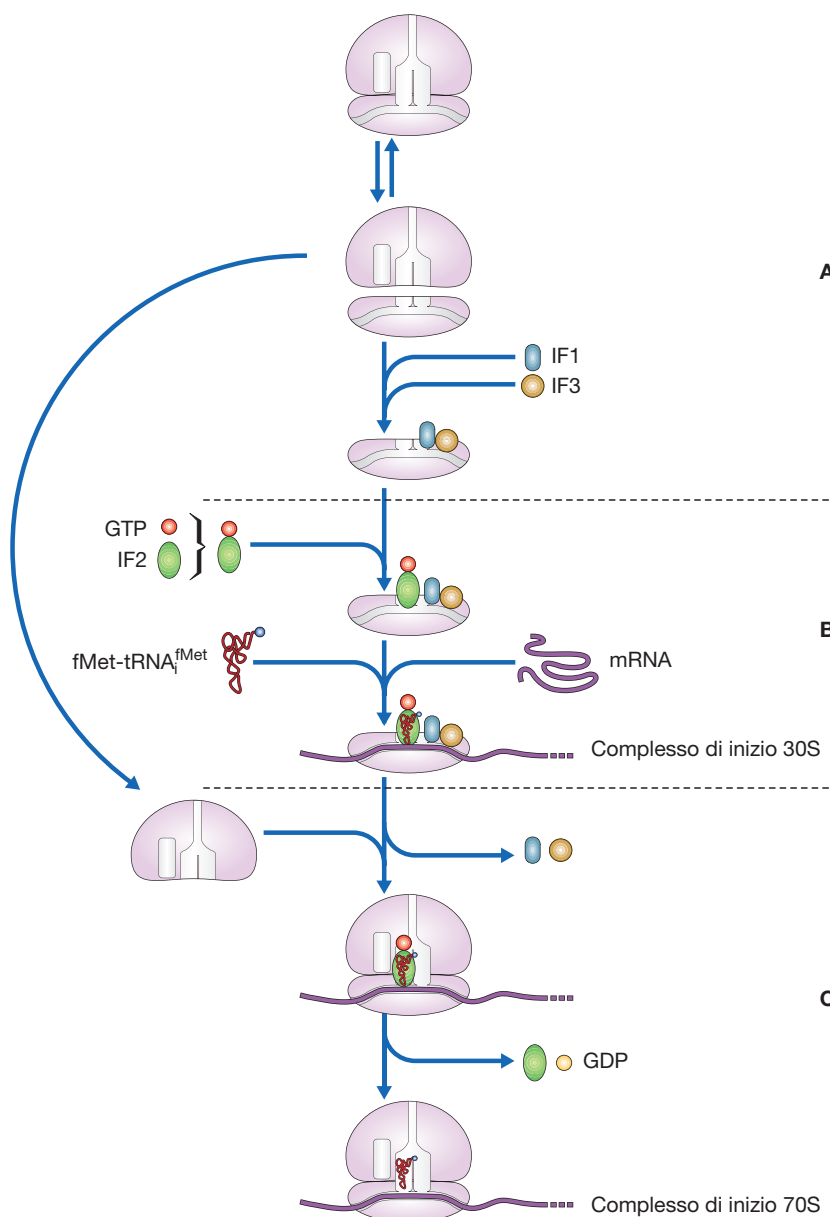
Processo di inizio della traduzione nei procarioti

La sequenza degli eventi descritta in **figura 15.6** è uno dei possibili modelli. Infatti, c'è ancora qualche incertezza sull'ordine temporale delle interazioni tra i vari componenti. Poiché dopo ogni ciclo di sintesi proteica i ribosomi vengono rilasciati in forma di ribosomi interi 70S, è necessario che questi vengano dissociati nelle due subunità affinché queste possano essere ri-utilizzate in un altro ciclo di sintesi. I ribosomi 70S tendono a dissociarsi fino a che viene raggiunto un equilibrio dinamico tra ribosomi interi e subunità e sicuramente i fattori IF1 e IF3 sono coinvolti in questa dissociazione, anche se non c'è ancora un'opinione concorde su alcuni dettagli della reazione.

Il modello mostrato in **figura 15.6A** implica un primo intervento di IF1, che legandosi alla 30S nel ribosoma intero contribuisce alla dissociazione delle due subunità. IF3, invece, interagisce solo con la subunità già dissociata e, impedendone la riassociazione con la subunità 50S, la sottrae all'equilibrio e la mantiene disponibile per un nuovo inizio. Dopo che IF1 e IF3 hanno portato a termine la dissociazione dei ribosomi nelle due subunità, la 30S può essere utilizzata per la formazione di un nuovo complesso d'inizio 30S che comprenderà, oltre alla subunità stessa, tre fattori e un tRNA di inizio. Come mostrato in **figura 15.6B**, IF2-GTP va a legarsi alla subunità 30S in prossimità dell'emisito P, dove promuove l'ingresso del tRNA di inizio $\text{fMet-tRNA}_i^{\text{fMet}}$. Questo è uno dei passaggi che presenta qualche incertezza; infatti, mentre in genere si pensa che IF2 interagisca prima con la subunità ribosomale per poi reclutare $\text{fMet-tRNA}_i^{\text{fMet}}$, altri suggeriscono che IF2 si leghi prima a $\text{fMet-tRNA}_i^{\text{fMet}}$ e poi lo trasporti nel sito P, analogamente a quello che succede durante l'allungamento, dove il fattore EF-Tu lega un amminoacil-tRNA e lo porta nel sito A (*vedi* oltre). È interessante notare che il tRNA di inizio, $\text{fMet-tRNA}_i^{\text{fMet}}$, è l'unico tRNA che entra nel ribosoma senza passare attraverso il sito A, come invece fanno tutti gli altri tRNA portati dal fattore di allungamento EF-Tu. Comunque, contemporaneamente o subito dopo il posizionamento di $\text{fMet-tRNA}_i^{\text{fMet}}$ nell'emisito P si ha il legame, mediato da IF3, dell'mRNA alla subunità 30S. Come detto precedentemente, questa interazione è facilitata e resa specifica dall'appaiamento di basi complementari tra la sequenza di Shine-Dalgarno e l'estremità 3' dell'rRNA 16S nella subunità piccola del ribosoma e tra il codone AUG e l'anticodone del tRNA di inizio $\text{fMet-tRNA}_i^{\text{fMet}}$ già posizionato nell'emisito P. È da notare che nel complesso di inizio 30S la presenza di IF1 e IF2, rispettivamente nelle adiacenze degli emisiti A e P, impedisce l'interazione della subunità ribosomale con qualsiasi altro tRNA carico, tranne l' $\text{fMet-tRNA}_i^{\text{fMet}}$ specificamente posizionato da IF2. Un'altra interessante osservazione è che se si va a saggiare individualmente l'interazione dei tre fattori con la subunità 30S, si trova che si formano legami poco stabili, mentre l'interazione si fa molto più stabile se si fa interagire la 30S contemporaneamente con i tre fattori.

Il passo successivo è l'associazione del complesso di inizio 30S con la subunità ribosomale 50S, che risulta nella formazione del complesso di inizio 70S. Come mostrato in **figura 15.6C**, perché ciò avvenga è necessario, dapprima, il rilascio

Figura 15.6 Inizio della traduzione nei procarioti. (A) I fattori IF1 e IF3 legano la subunità 30S libera e impediscono che si riassoci alla 50S, mantenendola così disponibile per il legame all'mRNA. (B) La subunità 30S con IF3 si lega all'mRNA sul sito di inizio (riconosciuto mediante appaiamento di basi tra l'estremità 3' dell'rRNA 16S e la sequenza di Shine-Dalgarno). In corrispondenza dell'emisita P si lega anche IF2-GTP, a cui poi si unisce il tRNA iniziatore $fMet-tRNA_i^{fMet}$, completando così la formazione del cosiddetto complesso di inizio 30S. (C) Il rilascio dei fattori IF1 e IF3 permette il legame della subunità 50S; il successivo rilascio anche di IF2 con idrolisi del GTP porta al complesso di inizio 70S.



dal complesso 30S dei fattori IF1 e IF3, che permette l'assemblaggio della subunità grande e, successivamente, il rilascio di IF2, che dipende dall'idrolisi del GTP a GDP. Il complesso d'inizio 70S, costituito da un ribosoma intero 70S, un mRNA e un $fMet-tRNA_i^{fMet}$, è pronto per entrare nella fase di allungamento della traduzione.

15.2 Inizio negli eucarioti

Come nei procarioti, anche negli eucarioti la traduzione inizia con la formazione di un complesso tra la subunità ribosomale minore, l'mRNA e il tRNA di inizio, che vanno a costituire il cosiddetto **complesso di pre-inizio 43S**. Tuttavia, l'inizio della sintesi proteica è la fase del processo che più differisce tra procarioti ed eucarioti. Innanzitutto, anche se, come nei procarioti, negli eucarioti abbiamo

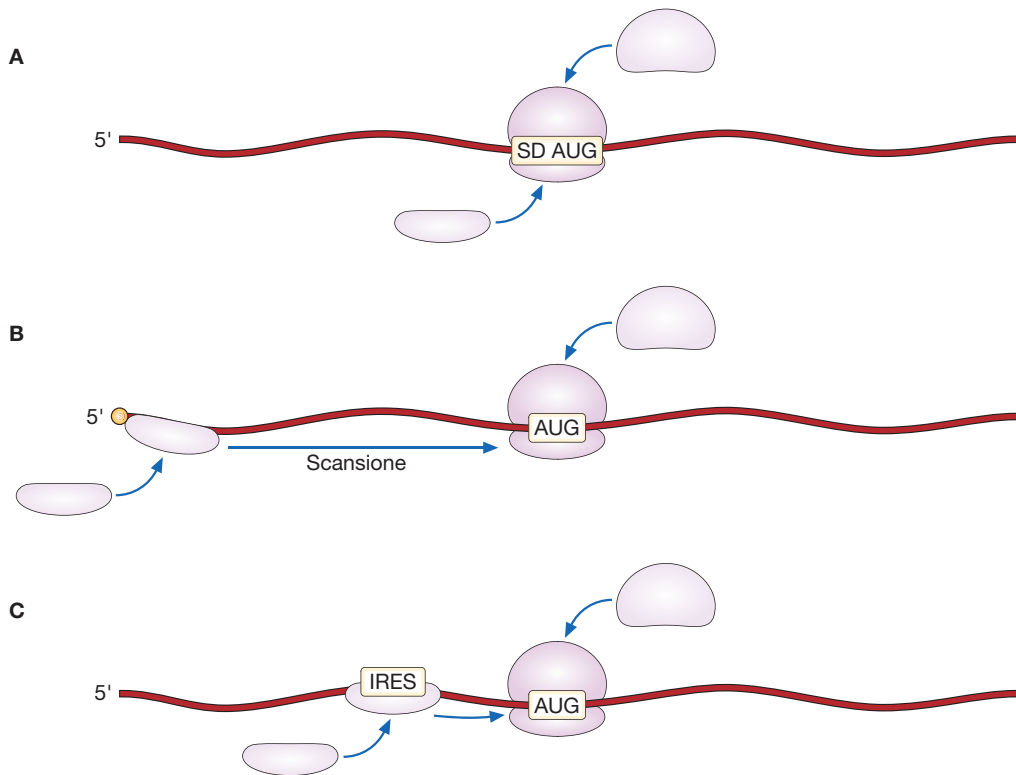


Figura 15.7 Confronto schematico dei meccanismi di inizio della traduzione. (A) Nei procarioti si ha un inizio interno dipendente dalla presenza della sequenza di Shine-Dalgarno (SD) poco a monte dell'AUG di inizio. (B) Negli eucarioti in genere la subunità piccola lega il cap all'estremità 5' dell'mRNA e poi procede nella

scansione della 3'UTR fino a trovare l'AUG di inizio. (C) Per alcuni virus e mRNA eucariotici si ha un meccanismo di inizio cap-indipendente in cui la subunità piccola lega una sequenza IRES interna e da lì procede per scansione fino all'AUG di inizio.

una metionina di inizio, questa non è formilata. Esistono comunque due tRNA per la metionina, uno utilizzato per l'inizio, chiamato $\text{tRNA}_i^{\text{Met}}$, e uno per la fase di allungamento, tRNA^{Met} . Un'altra differenza sostanziale, schematizzata in **figura 15.7A e B**, è il meccanismo di interazione tra mRNA e subunità piccola del ribosoma che porta alla formazione del complesso di inizio. Nei procarioti, come schematizzato in **figura 15.7A**, la presenza della sequenza di Shine-Dalgarno permette un inizio della traduzione "interno" all'mRNA, e quindi anche la possibilità di mRNA policistronici. Negli eucarioti, invece, la subunità minore del ribosoma interagisce inizialmente con l'estremità 5' dell'mRNA e da lì migra, mediante un processo di "scansione" (*scanning*), lungo la 5'UTR fino a trovare il primo AUG, che verrà utilizzato come codone di inizio della traduzione (**fig. 15.7B**). È chiaro che questo meccanismo di inizio è incompatibile con la presenza di più sequenze codificanti in una singola molecola di mRNA; infatti la grande maggioranza degli mRNA eucariotici è monocistronica.

Fattori di inizio negli eucarioti

Mentre, come abbiamo visto sopra, per i procarioti sono necessari tre fattori, ciascuno costituito da una singola catena polipeptidica, negli eucarioti i fattori sono più del doppio e, soprattutto, alcuni di essi sono molto grandi e costituiti da numerose subunità, per cui in totale sono necessari oltre 30 diversi polipeptidi per l'inizio della traduzione. Per dare un'idea della complessità del sistema, la **tabella 15.1** riporta l'elenco dei fattori di inizio della traduzione nei mammiferi

Tabella 15.1 Fattori di inizio per la traduzione cap-dipendente degli eucarioti.

Fattore	Ruolo
eIF1	Lega la subunità 40S, controlla l'accuratezza del riconoscimento del codone AUG di inizio e contribuisce con eIF3 a impedire l'associazione prematura con la 60S
eIF1A	Promuove il legame di Met-tRNA _i ^{Met} alla subunità 40S e contribuisce con eIF3 all'interazione forte di questa con l'mRNA
eIF2 (3 subunità)	Interagisce con Met-tRNA _i ^{Met} formando il complesso ternario "Met-tRNA _i ^{Met} /eIF2/GTP"
eIF2B (5 subunità)	Catalizza lo scambio GTP/GDP per il riciclo di eIF-2 (è chiamato anche GEF = <i>Guanine-nucleotide Exchange Factor</i>)
eIF3 (12 subunità)	Lega la subunità 40S e ne promuove il legame con Met-tRNA _i ^{Met} e mRNA e impedisce l'associazione della 60S prima che si sia formato il complesso di inizio 43S
eIF4E	Riconosce e lega il 5'cap
eIF4 eIF4F eIF4G	Costituisce un'impalcatura per l'assemblaggio di eIF4F, interagendo con eIF4E ed eIF4A
eIF4A	Ha attività RNA elicasi ATPasi-dipendente
eIF4B	Stimola l'elicasa e lega simultaneamente eIF4F
eIF5	Interagisce con il complesso di pre-inizio 43S (40S/Met-tRNA _i ^{Met} /eIF3/eIF2-GTP) e promuove l'idrolisi del GTP associato a eIF2
eIF5B	Legato a GTP recluta il complesso ternario "Met-tRNA _i ^{Met} /eIF2/GTP" sulla subunità 40S
eIF6	Lega la subunità 60S libera e impedisce la sua associazione con la 40S

(*H. sapiens*), anche se qui ci limiteremo a considerare solo i principali. I fattori d'inizio eucariotici vengono indicati con la sigla **eIF** (*eukaryotic Initiation Factor*) cui seguono numeri e lettere di specificazione.

Processo di inizio della traduzione negli eucarioti

La **figura 15.8** mostra schematicamente la sequenza degli eventi più importanti che portano all'assemblaggio del complesso di inizio negli eucarioti, anche se esistono notevoli incertezze sull'ordine temporale di alcune interazioni. La **figura 15.8A** mostra la formazione del cosiddetto **complesso di pre-inizio 43S**. I fattori eIF3 ed eIF1 svolgono essenzialmente le stesse funzioni di IF3 e IF1 procariotici, cioè contribuiscono alla dissociazione del ribosoma intero 80S nelle due subunità legando la subunità piccola 40S e alla stabilizzazione del complesso d'inizio. Il reclutamento del tRNA di inizio sulla 40S è mediato da due proteine: eIF2-GTP ed eIF5. La prima interagisce con il tRNA di inizio Met-tRNA_i^{Met} formando un complesso ternario (Met-tRNA_i^{Met}/eIF2/GTP) il cui reclutamento sulla 40S necessita di eIF5. Si ha così la formazione del complesso di pre-inizio 43S che comprende la subunità 40S, il Met-tRNA_i^{Met} e i fattori fino a qui riportati. Nel frattempo e indipendentemente si forma un complesso tra l'mRNA e il fattore eIF4 (**fig. 15.8B**). Questo è composto dalle subunità eIF4A, eIF4B, eIF4E, eIF4G. Quest'ultima è una grande proteina (220 kDa) che costituisce un'impalcatura su cui si monta il complesso. Come mostrato in figura, essa è legata a eIF4E e a eIF4A, insieme alle quali costituisce l'intermedio che prende il nome di **eIF4F**. Questo interagisce con l'mRNA mediante un'interazione specifica tra la subunità eIF4E e il cap (cappuccio) che si trova legato al primo nucleotide di tutti gli mRNA eucariotici. Infatti eIF4E viene anche chiamata **cap binding protein**. Al complesso si aggiunge poi la subunità eIF4B.

Successivamente, l'interazione tra il complesso di pre-inizio 43S e il complesso eIF4F/mRNA risulta nella formazione del complesso di inizio 48S in cui l'estre-

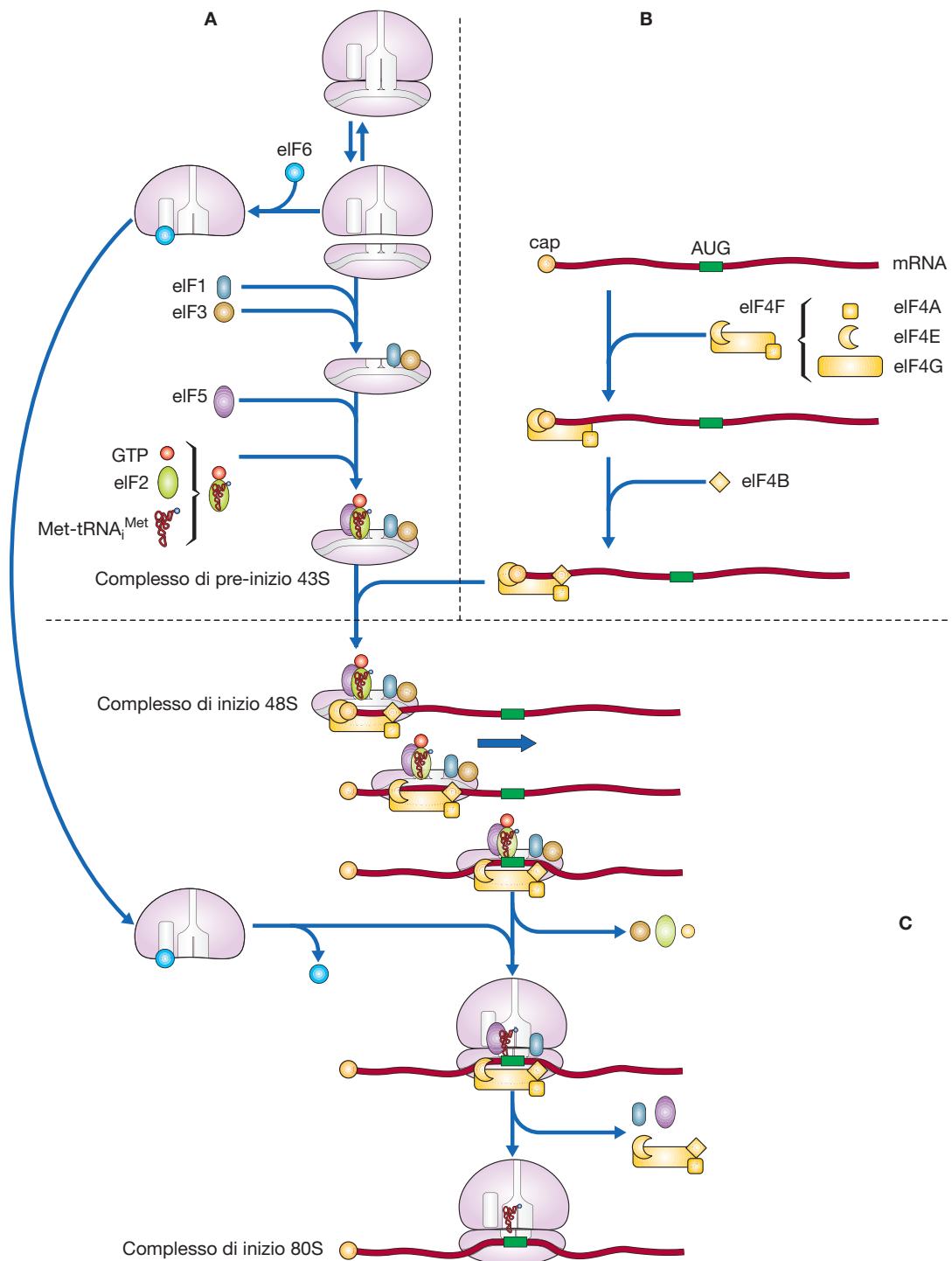


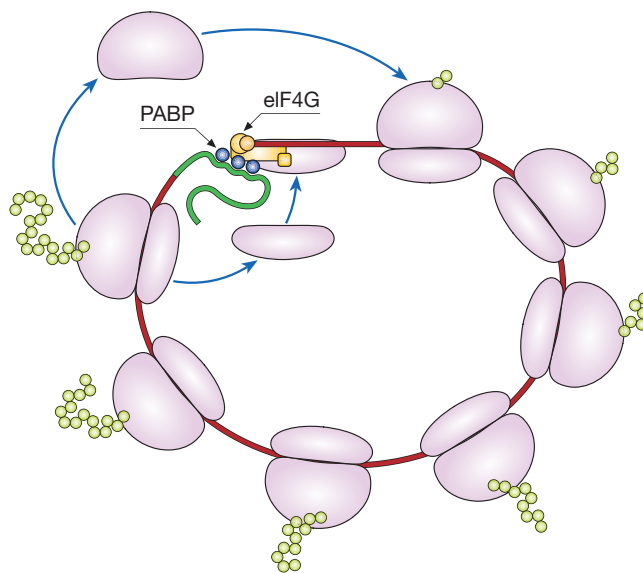
Figura 15.8 Inizio della traduzione cap-dipendente negli eucarioti. (A) Formazione del complesso di pre-inizio 43S costituito dalla subunità piccola del ribosoma complessata con i fattori eIF1, eIF3, eIF5-GTP e il complesso ternario Met-tRNA_i^{Met}/eIF2/GTP. (B) Formazione del complesso tra mRNA e il fattore eIF4. (C) L'interazione tra il complesso di pre-inizio 43S e il complesso eIF4/mRNA risulta nella formazione del com-

plesso di inizio 48S in cui l'estremità 5' dell'mRNA è legata alla subunità ribosomiale 40S. Successivamente, il processo di scansione porta la subunità ribosomiale 40S con i fattori associati fino al codone AUG di inizio. L'unione della subunità ribosomiale 60S e il rilascio di tutti i fattori portano alla formazione del complesso di inizio 80S.

mità 5' dell'mRNA è legata alla subunità ribosomale 40S (**fig. 15.8C**). A questo punto la subunità 40S con i fattori a essa associati inizia il processo di scansione, ATP-dipendente, lungo la 5'UTR fino a trovare il codone di inizio AUG. Poiché la scansione può essere ostacolata dalla presenza di strutture secondarie che si formano per appaiamento intramolecolare tra sequenze di basi complementari nella 5'UTR dell'mRNA, è necessaria un'attività elicasica che sciogla queste strutture. Tale attività è fornita dai fattori **eIF4A** ed **eIF4B**. eIF4A è una RNA elicasi della famiglia delle proteine DEAD, così chiamate perché contengono la sequenza di consenso Asp (D) Glu (E) Ala (A) Asp (D), mentre eIF4B ha una funzione di stimolazione su eIF4A. Il modello più semplice per la scansione prevede che la subunità riconosca il primo AUG che incontra e da lì inizi la traduzione. A questa regola esistono però delle eccezioni. Infatti, per una piccola percentuale di mRNA (circa il 5%) la subunità salta uno o più AUG per andare a iniziare su uno più a valle. Questo fa pensare che ci sia qualcosa che permetta di distinguere il vero AUG di inizio. Un'analisi attenta delle sequenze che circondano gli AUG di inizio ha permesso di identificare un consenso, anche se non molto stringente. Questo consenso (RNNAUGG, dove R indica una purina) viene spesso chiamato **sequenza di Kozak** dal nome di Marilyn Kozak, la ricercatrice che l'ha individuata e studiata. Durante la scansione la subunità 40S porta, nell'emisito P, il tRNA iniziatore Met-tRNA^{Met} il cui anticodone va ad appaiarsi con l'AUG sull'mRNA contribuendo così al riconoscimento del sito di inizio della traduzione. Finalmente, quando la subunità 40S è posizionata sul codone di inizio AUG, può essere reclutata la subunità ribosomale 60S formando così il complesso di inizio 80S. Come mostrato in **figura 15.8A e C**, perché ciò avvenga è necessario il fattore antiassociativo **eIF6**. Questo ha per la 60S una funzione analoga a eIF3 per la 40S, cioè lega la 60S impedendo la sua riassociazione con la subunità 40S mantenendola così disponibile per la sua utilizzazione. Affinché la 60S possa essere reclutata, è anche necessario che il complesso di inizio 48S rilasci i fattori eIF2 e eIF3. Infine, come avviene nei procarioti, il reclutamento della subunità 60S porta al rilascio dei rimanenti fattori di inizio, con idrolisi di GTP da parte di eIF5B, analogo del procariotico IF2.

Per non complicare troppo la **figura 15.8** e la descrizione dell'inizio della traduzione negli eucarioti abbiamo finora considerato l'mRNA come una molecola più o meno lineare che viene letta dal ribosoma dall'estremità 5' a sinistra

Figura 15.9 Modello della “circularizzazione” dell'mRNA eucariotico. La proteina (PABP) associata in più copie alla coda di poli(A) interagisce con il fattore di inizio eIF4G avvicinando così le due estremità dell'mRNA. La struttura circolare assunta dall'mRNA facilita l'inizio della traduzione da parte dei ribosomi, che, terminando la traduzione al 3' dell'mRNA, vengono rilasciati in prossimità del 5' e quindi rapidamente reclutati per un nuovo ciclo di traduzione.



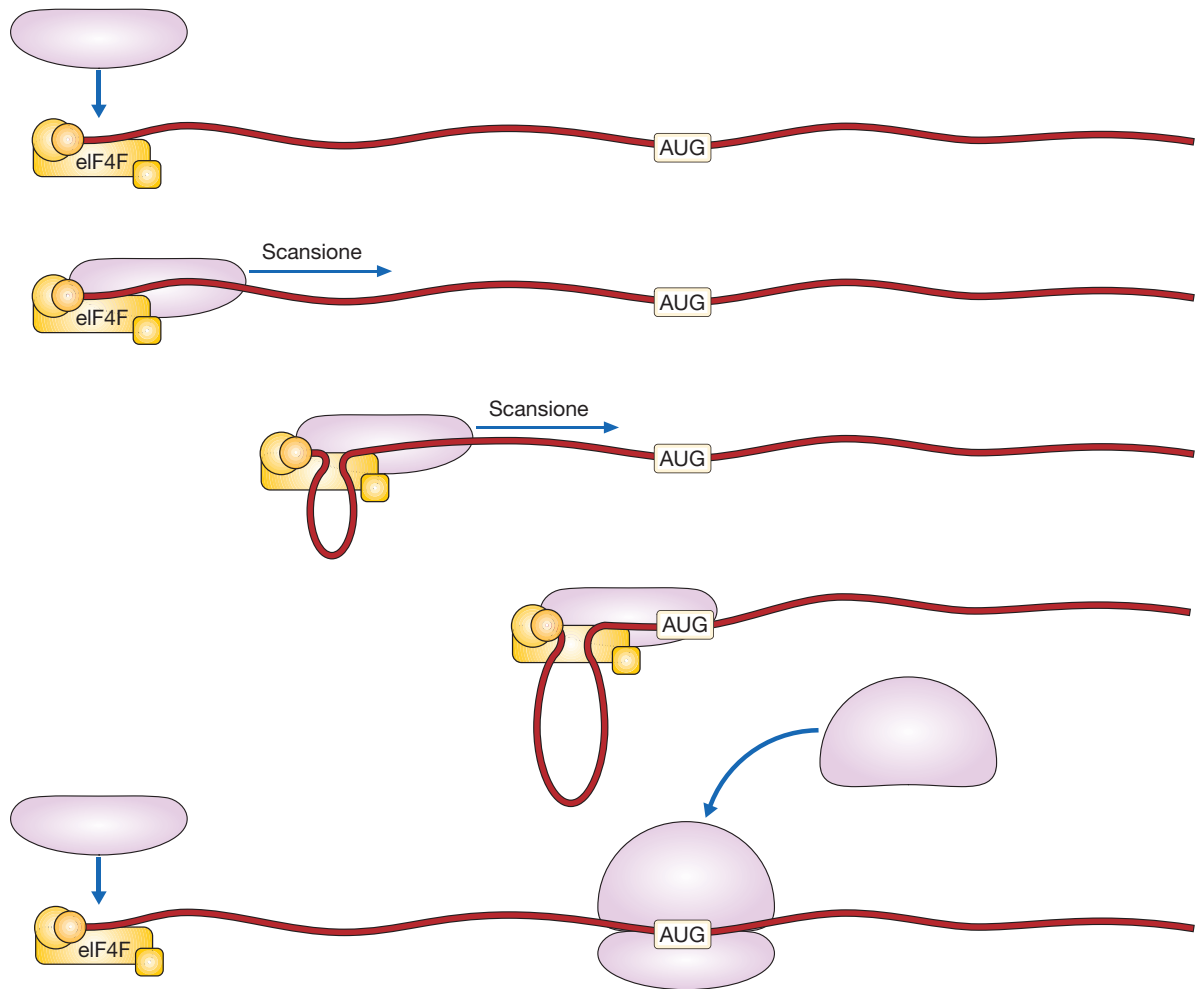


Figura 15.10 Durante la scansione il fattore eIF4F, legato alla subunità 40S, mantiene il legame con il 5'cap. La contemporanea interazione di eIF4F con il cap e con la subunità 40S fa sì che la regione della 5'UTR già scansionata formi un'ansa che va allungandosi fino a quando la subunità 40S raggiunge il codone

di inizio. A questo punto la subunità 40S lega una 60S e rilascia eIF4F, che può legare una nuova subunità 40S. Per semplicità è stato riportato solo il fattore di inizio eIF4F (con le sue subunità E, G e A) e sono stati omessi gli altri fattori.

verso l'estremità 3' a destra, e abbiamo omesso di menzionare un'altra proteina coinvolta nell'inizio della traduzione, la **PABP**. Questa è una proteina associata in varie copie alla coda poli(A) che si trova all'estremità 3' degli mRNA eucariotici. È stato trovato che la PABP, pur rimanendo legata al poli(A), interagisce con il fattore di inizio eIF4G associato all'estremità 5' dello stesso mRNA. Come mostrato in **figura 15.9**, questa interazione porta a contatto le due estremità dell'mRNA, che quindi assume una configurazione circolare (o ad ansa), cosicché i ribosomi che terminano la traduzione al 3' dell'mRNA vengono rilasciati in prossimità del 5'. Secondo un modello sostenuto da vari autori, la struttura ad ansa dell'mRNA non solo avrebbe un ruolo nella stabilizzazione dell'mRNA proteggendolo da attività esonucleasiche, ma ne aumenterebbe l'efficienza di traduzione.

Un'altra cosa che abbiamo fin qui omesso di menzionare è che, durante il processo di scansione della subunità 40S con tutti i fattori associati lungo la 5'UTR dell'mRNA, il fattore eIF4F resta legato al cap tramite la sua subunità eIF4E. Quindi, come mostrato in **figura 15.10**, la 5'UTR dell'mRNA forma un'ansa che si allunga mano a mano che il complesso avanza su di esso. Una volta che la

subunità 40S ha raggiunto l'AUG di inizio ed è stata reclutata la subunità 60S, tutti i fattori vengono rilasciati, ma eIF4F resterebbe legato al 5'cap dell'mRNA pronto a interagire con un altro complesso di pre-inizio 43S.

Meccanismo di inizio alternativo cap-indipendente negli eucarioti

La grande maggioranza degli mRNA eucariotici utilizza il meccanismo di inizio sopra descritto che, essendo basato sul riconoscimento del 5'cap, viene definito **inizio cap-dipendente**. Esistono però mRNA eucariotici che utilizzano meccanismi di inizio diversi. Il caso più notevole è quello dell'inizio interno, schematizzato in **figura 15.7C**, che è stato scoperto e studiato in alcuni virus che infettano le cellule eucariotiche, ma poi riscontrato anche per alcuni mRNA cellulari. Virus della famiglia dei picornavirus hanno il genoma costituito da una molecola di RNA che funge anche da mRNA. Questo RNA non è fornito di 5'cap e quindi non può sostenere una sintesi proteica che utilizzi un meccanismo cap-dipendente. Tuttavia, può essere tradotto perché contiene, nel tratto che precede la regione codificante, una particolare sequenza nucleotidica capace, con l'ausilio di specifici fattori proteici, di reclutare direttamente la subunità ribosomale 40S del ribosoma e operare così un inizio di traduzione "interno", analogamente a quanto abbiamo visto succedere nei procarioti. Le sequenze di RNA che hanno la proprietà di reclutare la subunità 40S indipendentemente dal 5'cap vengono chiamate **IRES** (*Internal Ribosome Entry Site*, sito interno di entrata del ribosoma). Perché questi virus utilizzano questo particolare meccanismo di inizio della traduzione? Ciò si spiega col fatto che questi stessi virus, quando infettano la cellula eucariotica, producono un enzima che taglia il fattore eIF4G, inibendo così la traduzione cap-dipendente degli mRNA endogeni della cellula ospite e sostenendo la sintesi delle proprie proteine utilizzando un meccanismo di inizio cap-indipendente.

Più recentemente sono stati osservati alcuni rari casi di mRNA cellulari caratterizzati da una regione 5'UTR molto estesa che pure hanno la capacità di un simile "inizio cap-indipendente", ma il significato funzionale di questo diverso meccanismo d'inizio non è chiaro.

15.3 Allungamento

Abbiamo visto che al termine del processo di inizio si ha un ribosoma intero legato all'mRNA in corrispondenza del codone di inizio AUG. A questo codone è appaiato l'anticodone del tRNA di inizio che si trova posizionato nel sito P del ribosoma stesso, mentre il sito A è vuoto. Da questa situazione si procede alla fase di allungamento della traduzione, che consiste in una serie di cicli, uno per ogni amminoacido da aggiungere. Ogni ciclo di allungamento è composto da tre passaggi (**fig. 15.11**):

1. l'entrata nel sito A del ribosoma di un nuovo tRNA carico;
2. il trasferimento del primo amminoacido o della catena peptidica nascente dal tRNA che si trova nel sito P sul nuovo amminoacido che si trova legato al tRNA posizionato nel sito A, e che comporta la formazione del nuovo legame peptidico;
3. la traslocazione di tutto l'insieme (codone dell'mRNA/tRNA/peptide nascente) dal sito A al sito P, in modo che l'mRNA esponga sul sito A, ormai vuoto, il codone seguente. Si ristabiliscono così le condizioni iniziali, con il sito P occupato dal peptidil-tRNA e il sito A vuoto, da cui si procede per un ulteriore ciclo.

L'intero processo è evolutivamente molto conservato, quindi molto simile in procarioti ed eucarioti.

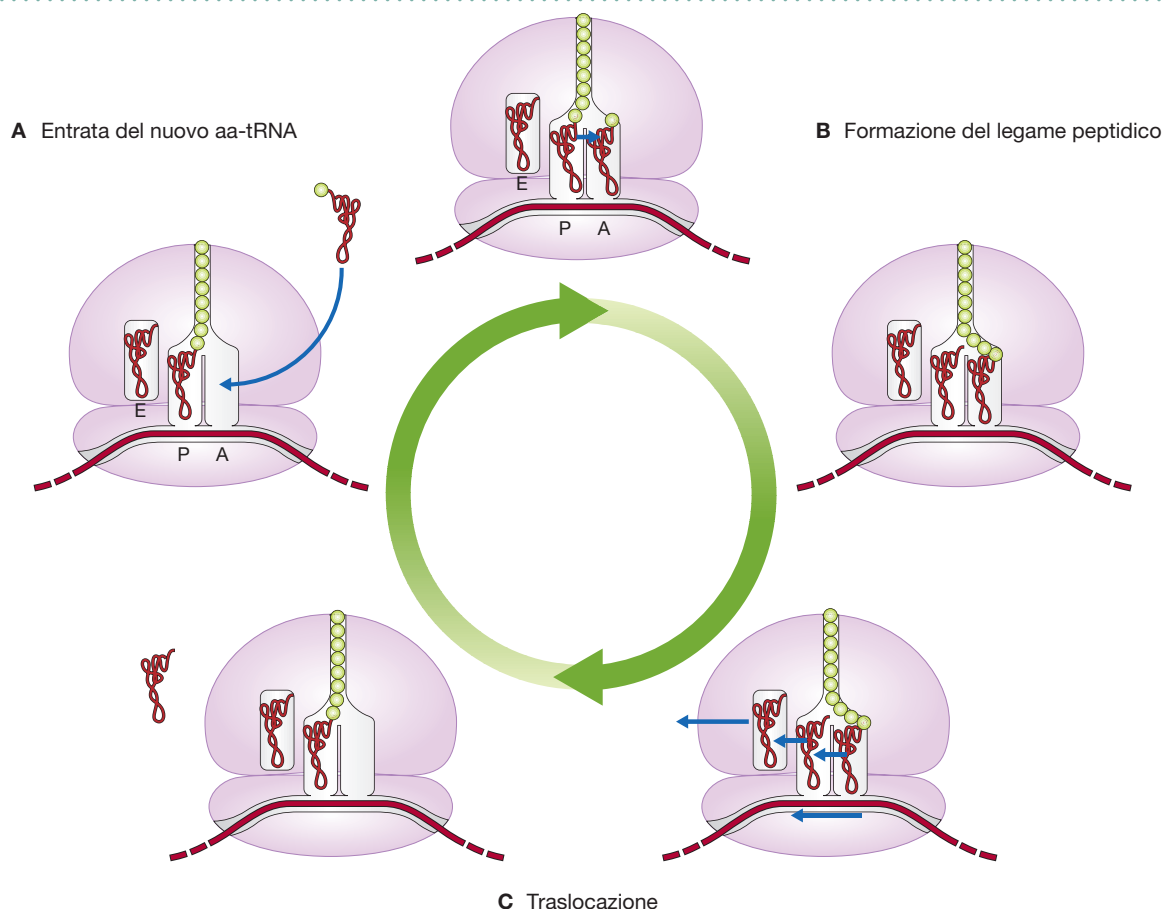


Figura 15.11 Rappresentazione schematica del ciclo di allungamento della traduzione. (A) Un nuovo aa-tRNA entra nel sito A, mentre i siti E e P sono occupati, rispettivamente, dal tRNA scarico e dal tRNA che porta il peptide nascente (peptidil-tRNA). (B) Formazione del legame peptidico: l'attività peptidil transferasica del ribosoma catalizza il trasferimento del peptide nascente dal tRNA del sito P all'amminoacido dell'aa-tRNA nel sito A. (C) Traslocazio-

ne: spostamento del ribosoma rispetto ai tRNA e all'mRNA di un codone (3 nucleotidi). Il tRNA scarico che era già nel sito E viene rilasciato, il tRNA appena scaricato che si trova nel sito P passa nel sito E e il tRNA che ora porta la catena peptidica nascente si sposta dal sito A al sito P; contemporaneamente si sposta l'mRNA legato con due codoni a questi ultimi due tRNA.

Nei procarioti, per il processo di allungamento sono necessari, oltre all'mRNA, ai ribosomi e a tutti i diversi tRNA caricati con i corrispondenti amminoacidi, anche tre fattori proteici, **EF-Tu**, **EF-Ts** ed **EF-G** (**EF**, *Elongation Factor*), e inoltre GTP (**fig. 15.12**). All'inizio di ogni ciclo di allungamento il sito P del ribosoma è occupato dall'ultimo tRNA che è entrato e che porta alla sua estremità 3' la catena peptidica nascente (o fMet-tRNA^{fMet} se si tratta del primo ciclo di allungamento), mentre il sito A è vuoto. Il sito E del ribosoma è occupato dal precedente tRNA ormai scarico oppure è vuoto se nel sito P abbiamo fMet-tRNA^{fMet}. Durante la prima fase del ciclo entra nel sito A un nuovo amminoacil-tRNA, che può essere uno qualunque degli amminoacil-tRNA, tranne fMet-tRNA^{fMet}, che è l'unico a entrare direttamente nel sito P. L'ingresso dell'amminoacil-tRNA nel sito A è mediato dal fattore EF-Tu. Questo lega dapprima una molecola di GTP formando un complesso binario EF-Tu-GTP. Il legame poi con un amminoacil-tRNA risulta nella formazione di un complesso ternario che è attivo nel trasferimento del tRNA carico al sito A del ribosoma. Per una corretta sintesi proteica è essenziale che in questo passaggio venga selezionato il giusto amminoacil-tRNA. L'accuratezza di questa selezione si basa sulle interazioni tra le basi del codone sull'mRNA e

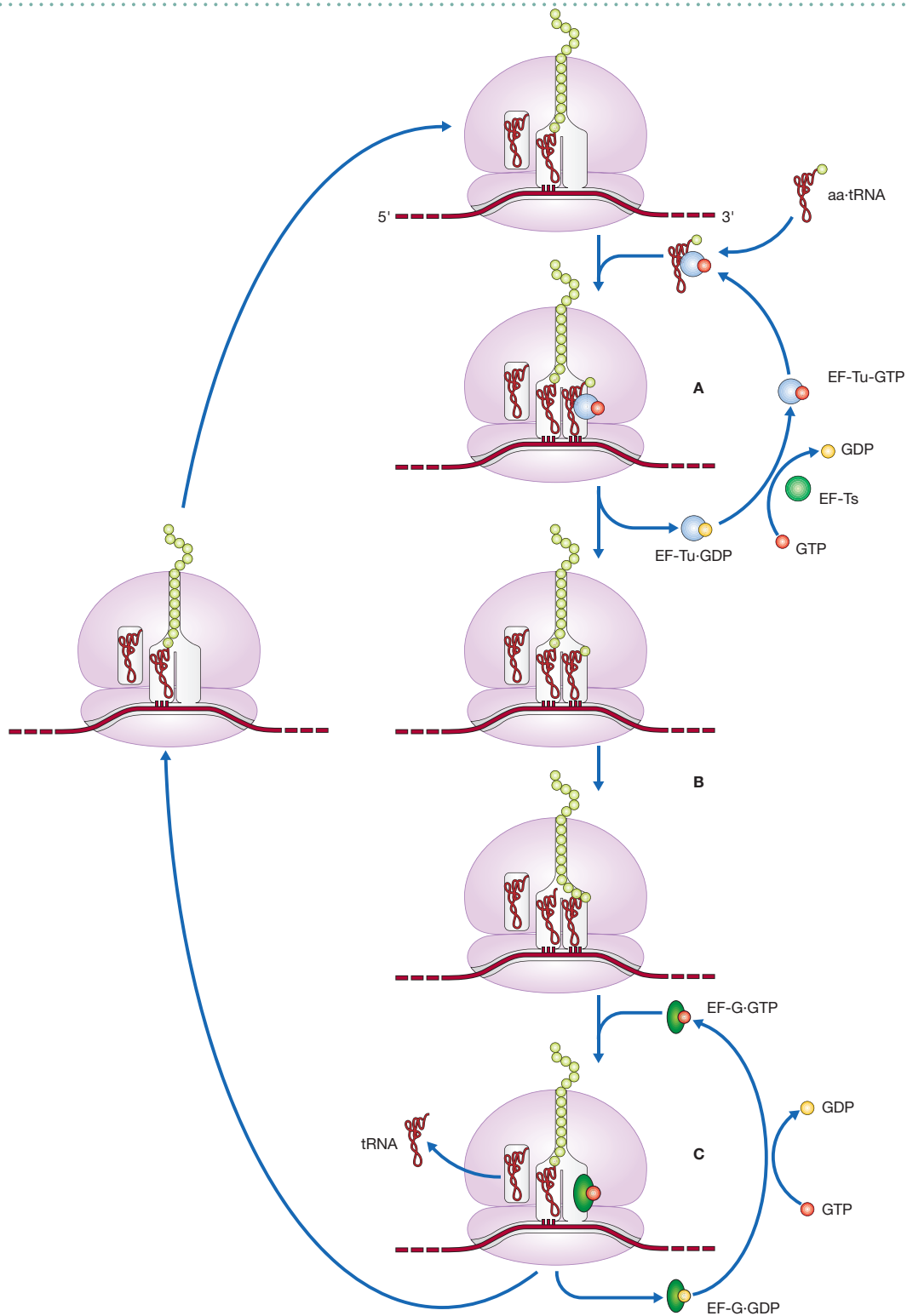


Figura 15.12 La fase di allungamento della traduzione nei procarioti. (A) Il complesso EF-Tu-GTP lega un aa-tRNA e lo porta nel sito A vuoto del ribosoma. Solo se l'appaiamento codone-anticodone è corretto avviene l'idrolisi del GTP in GDP e il complesso EF-Tu-GDP viene rilasciato per poi essere riattivato a EF-Tu-GTP a opera di EF-Ts. (B) Il peptide nascente che si trova sul tRNA nel sito P del ribosoma viene trasferito, dall'attività peptidil transfera-

sica del ribosoma, sull'amminoacido portato dal tRNA nel sito A. (C) La traslocazione dell'insieme peptidil-tRNA/mRNA dal sito A al sito P richiede il fattore EF-G associato a una molecola di GTP che verrà idrolizzata a GDP; contemporaneamente il tRNA scarico del sito P passa nel sito E, mentre quello del sito E viene rilasciato. Il fattore, rilasciato in forma di EF-G-GDP, viene poi riattivato a EF-G-GTP per potere essere riusato in un altro ciclo.

quelle dell'anticodone sul tRNA, secondo le regole del codice genetico. Queste interazioni avvengono nel "centro di decodificazione" della subunità minore del ribosoma che è costituito essenzialmente da regioni dell'rRNA 16S e che non contiene proteine ribosomali. Solo se l'appaiamento codone-anticodone è corretto, il ribosoma subisce un cambiamento conformazionale che porta all'idrolisi del GTP da parte dell'EF-Tu a cui esso è legato. A questa idrolisi segue il rilascio di EF-Tu-GDP e l'associazione stabile dell'amminoacil-tRNA al sito A del ribosoma. Il rilascio di EF-Tu-GDP è necessario perché il ribosoma possa procedere nella formazione del legame peptidico. Torneremo più avanti su questo aspetto della accuratezza della traduzione e sui meccanismi che la garantiscono. Una volta rilasciato, EF-Tu-GDP non è riutilizzabile per un altro ciclo se prima non viene riconvertito in EF-Tu-GTP. Come mostrato in **figura 15.12A**, la riattivazione di EF-Tu-GDP a EF-Tu-GTP è operata da un altro fattore di allungamento chiamato EF-Ts che agisce come fattore di scambio per il GTP.

È interessante notare che EF-Tu è la proteina più rappresentata in una cellula di *Escherichia coli*, circa il 5% delle proteine totali. Infatti, come già detto, in una cellula di *E. coli* ci sono circa 20 000 ribosomi, ognuno dei quali è circondato da una nuvola dei diversi tRNA per un totale di oltre 200 000 per cellula, ciascuno complessato con una molecola di EF-Tu.

Come mostrato schematicamente in **figura 15.12B**, la seconda fase del ciclo di allungamento consiste nella formazione del legame peptidico tra l'ultimo amminoacido della catena peptidica nascente, che si trova legato all'estremità 3' del tRNA presente nel sito P, e l'amminoacido portato dal nuovo amminoacil-tRNA appena entrato nel sito A, con conseguente spostamento della catena peptidica dal tRNA che si trova nel sito P al tRNA nel sito A. La reazione chimica è stata già descritta precedentemente in dettaglio e mostrata in **figura 15.2**. Chi catalizza la formazione del legame peptidico, che, in fin dei conti, rappresenta la vera essenza della sintesi proteica? Dopo ricerche infruttuose di uno specifico enzima o di un qualche fattore proteico accessorio che fosse coinvolto in questa reazione, è stato ipotizzato e poi dimostrato sperimentalmente che in effetti non esiste un enzima o fattore proteico di questo genere. Infatti, come abbiamo discusso nel paragrafo 14.1 l'attività peptidil transferasica è una proprietà del ribosoma stesso localizzata, in particolare, nella subunità grande in quello che abbiamo chiamato centro della peptidil transferasi, che è formato essenzialmente dall'rRNA 23S in una regione dove non sono presenti proteine ribosomali.

Il completamento di un ciclo di allungamento richiede una terza fase, chiamata **traslocazione**, che ristabilisce la situazione iniziale da cui procedere per il ciclo seguente. Come si osserva in **figura 15.12C**, durante la traslocazione, che possiamo considerare come spostamento del ribosoma di tre nucleotidi lungo la sequenza dell'mRNA o, se si preferisce, dell'mRNA rispetto al ribosoma, l'anticodone del peptidil-tRNA e il corrispondente codone sull'mRNA restano appaiati, cosicché si ha il loro contemporaneo spostamento dal sito A al sito P. Allo stesso tempo, il precedente tRNA ormai scarico si sposta dal sito P al sito di uscita E, da cui viene poi rilasciato. Alla fine della traslocazione l'mRNA espone il prossimo codone nel centro di decodificazione del sito A vuoto, che è così disponibile per iniziare un nuovo ciclo di allungamento. La traslocazione richiede il fattore di allungamento **EF-G**, la cui funzione è anche associata a idrolisi di GTP. Infatti, EF-G può associarsi al ribosoma solo se associato a GTP. L'interazione con il ribosoma del complesso EF-G-GTP inizia nell'emisito A della subunità grande e porta all'idrolisi del GTP. Successivamente, EF-G-GDP estende i suoi contatti all'emisito A della subunità minore del ribosoma promuovendo così la traslocazione del peptidil-tRNA dal sito A al sito P e del tRNA scarico

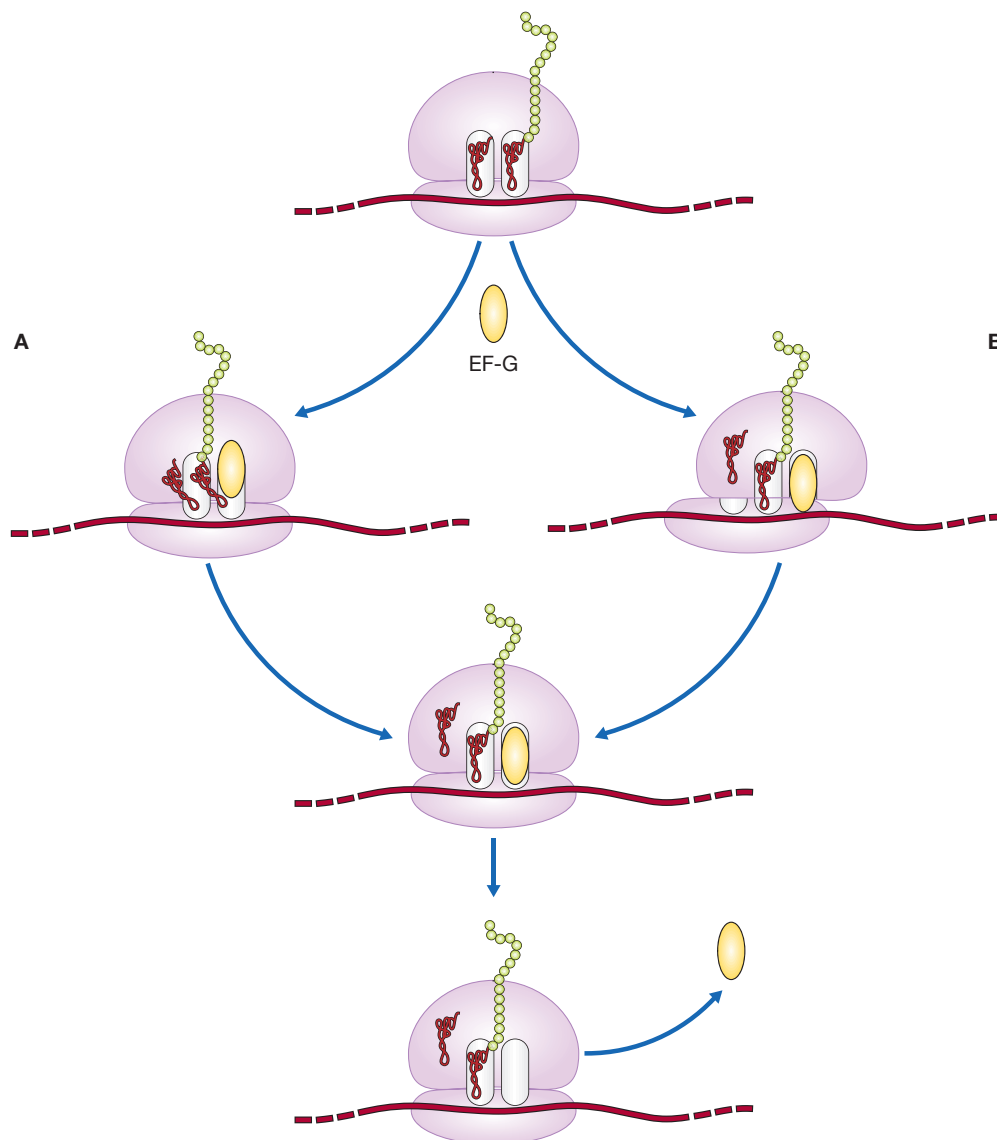


Figura 15.13 Due possibili modelli per spiegare il fatto che la traslocazione avviene prima nella subunità grande e poi in quella piccola. (A) Il peptidil-tRNA si sposta dal sito A al sito P in due tempi: dapprima si sposta l'estremità che porta il peptide nascente e successivamente si sposta l'estremità dell'anticodone. (B) Le due subunità del ribosoma si muovono una

rispetto all'altra formando per breve tempo un sito ibrido costituito dall'emisito A della subunità piccola e dall'emisito P della subunità grande, nel quale viene a trovarsi transitoriamente il peptidil-tRNA. Successivamente il ribosoma riprende la normale conformazione. [Fonte: Poole, E.S. et. al. (2007), *Cell Research*, 17, pp. 591-607.]

dal sito P al sito E, e l'avanzamento dell'mRNA di tre nucleotidi. Per potere essere riutilizzato in un altro ciclo di allungamento EF-G-GDP, analogamente a EF-Tu-GDP, deve essere riattivato a EF-G-GTP. In questo caso, però, non è necessario l'intervento di un fattore specifico essendo l'affinità di EF-G per GDP minore di quella per GTP.

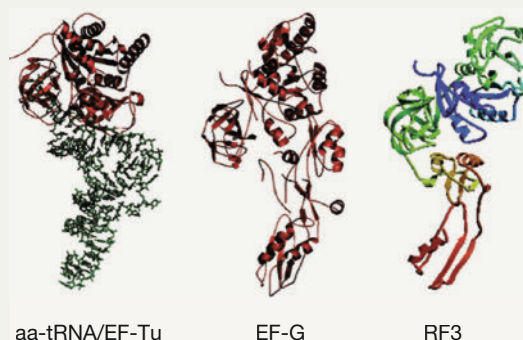
Abbiamo visto come EF-G interagisca prima con l'emisito A della subunità 50S e solo in un secondo tempo con l'emisito A della subunità 30S, provocando la traslocazione dei due tRNA in due fasi: prima si ha la traslocazione delle estremità accettrici dei tRNA dai siti P e A rispettivamente ai siti E e P; solo successivamente si ha la traslocazione delle altre estremità dei tRNA, quelle che contengono gli anticodoni, insieme all'mRNA. La **figura 15.13** mostra due possi-

FINESTRA 15.1

MIMETISMO MOLECOLARE DEI FATTORI CHE INTERAGISCONO CON IL SITO A DEL RIBOSOMA

Durante il susseguirsi dei cicli di allungamento della traduzione, il sito A del ribosoma viene occupato alternativamente dal fattore EF-Tu, che vi porta l'aa-tRNA, e dal fattore EF-G, che, dopo la formazione del legame peptidico, promuove la traslocazione. EF-Tu deve prima essere rilasciato perché si possa legare EF-G, e a sua volta EF-G deve essere rilasciato perché possa legarsi un nuovo complesso aa-tRNA/EF-Tu. Lo studio strutturale di questi componenti dell'apparato di traduzione ha messo in evidenza una loro straordinaria somiglianza strutturale spiegabile col fatto che devono interagire con lo stesso sito di legame nel ribosoma. Ma ciò che è più interessante e curioso è che, come mostrato in figura, questa somiglianza riguarda la struttura complessiva, che nel caso di EF-G è interamente proteica mentre nel caso di aa-tRNA/EF-Tu è in parte proteica e in parte costituita dal tRNA (grigio nella figura). In particolare, EF-G ha un dominio

proteico C-terminale, non presente in EF-Tu, che si presenta strutturalmente simile al tRNA legato a quest'ultimo. Questo fenomeno è stato chiamato **mimetismo molecolare**. Questa complessiva somiglianza strutturale non si limita ai due componenti sopra citati ma si estende ai fattori implicati nella terminazione della traduzione. Infatti, alla fine della fase di allungamento, quando il ribosoma raggiunge un codone di terminazione, viene innescata la fase di terminazione che comporta l'entrata nel sito A del fattore di rilascio RF1 (o di RF2), a cui poi si unisce RF3. Anche questo complesso RF1/RF3 (o RF2/RF3) ha una grande somiglianza strutturale con il complesso aa-tRNA/EF-Tu·GTP. Infatti RF1 e RF2 sono simili al dominio C-terminale di EF-G, che è quello che imita il tRNA presente in aa-tRNA/EF-Tu. Infine anche il fattore di riciclo dei ribosomi RRF ha una struttura che mima quella di un tRNA.



aa-tRNA/EF-Tu

EF-G

RF3

bili meccanismi per la traslocazione. Nel primo modello (**fig. 15.13A**), il ribosoma rimane strutturalmente fisso mentre i tRNA assumerebbero transitoriamente una disposizione obliqua rispetto al ribosoma in modo da trovarsi ciascuno con un'estremità in un sito e l'altra nell'altro sito. Il secondo modello, oggi il più accreditato, propone che a ogni ciclo di allungamento le due subunità ribosomali si spostino una rispetto all'altra (**fig. 15.13B**). Si ha così la formazione transitoria di un sito ibrido composto dall'emisito A della subunità piccola e dall'emisito P della subunità grande, in cui si troverebbe momentaneamente collocato il peptidil-tRNA. Con la traslocazione anche delle estremità degli anticodoni, il ribosoma riacquisterebbe la conformazione normale.

Il processo di allungamento della traduzione negli eucarioti è molto simile a quello descritto per i procarioti. I fattori di allungamento della traduzione eucariotica vengono chiamati **eEF** (*eukaryotic Elongation Factor*). Il fattore eEF1A è l'omologo di EF-Tu procariotico essendo responsabile dell'inserimento dell'amminoacil-tRNA nel sito A del ribosoma. L'attività di eEF1A dipende dal legame con una molecola di GTP e dall'idrolisi di questo in GDP. La forma attiva con GTP viene ripristinata ad opera del fattore eEF1B, che quindi è l'omologo del procariotico EF-Ts. Il fattore eEF2 è implicato nella fase di traslocazione, ed è quindi l'omologo dell'EF-G procariotico.

15.4 Terminazione

Il ciclo di allungamento si ripete numerose volte aggiungendo un amminoacido alla volta alla catena peptidica nascente fino a quando un codone di stop (UAG, UAA o UGA) dell'mRNA entra nel sito A del ribosoma. Questi codoni di stop (o terminazione, o non senso) sono chiamati anche amber (UAG), ochre (UAA) e opal (UGA), dal nome delle mutazioni che ne hanno permesso la scoperta (*vedi* finestra 3.1). In una prima fase (**fig. 15.14A**) la terminazione porta al rilascio della catena peptidica dall'ultimo tRNA e poi, in una seconda fase detta di post-terminazione (**fig. 15.14B**), si ha la dissociazione del tRNA, dell'mRNA e delle subunità ribosomali. Non esistono tRNA con anticodoni complementari ai codoni di stop. Questi sono invece riconosciuti da fattori proteici di terminazione **RF** (*Release Factor*, fattori di rilascio). Come mostrato in **figura 15.14**, nei procarioti i fattori RF1 e RF2 (detti di classe 1) entrano nel sito A del ribosoma riconoscendo direttamente i codoni di terminazione: **RF1** riconosce UAA e UAG, mentre **RF2** riconosce UGA e UAA. Essi inducono l'idrolisi del polipeptide dal peptidil-tRNA, o direttamente, oppure indirettamente provocando un cambiamento strutturale del ribosoma, che a sua volta idrolizza il peptidil-tRNA. In ogni caso, è necessario poi un fattore di classe 2 (**RF3**) che provoca il rilascio dal ribosoma del fattore di classe 1. Come abbiamo visto per altri fattori traduzionali, anche RF3 è una proteina che lega GTP. In realtà RF3 entra in scena legato a GDP interagendo, a livello del sito A ribosomale, con il fattore RF di classe 1 (RF1 o RF2) che vi si trova. In seguito alla liberazione del polipeptide stimolata dal fattore RF di classe 1, questo stimola lo scambio GDP-GTP su RF3, che a sua volta provoca il rilascio del fattore di classe 1 dal ribosoma. Questa prima fase della terminazione (**fig. 15.14A**) che ha portato al rilascio del polipeptide completo dal ribosoma, lascia però intatto il complesso ribosoma/mRNA/tRNA. Questo viene successivamente disassemblato per effetto del **fattore RRF** (*Ribosome Recycling Factor*, fattore di riciclaggio del ribosoma) e di EF-G (**fig. 15.14B**). Secondo un modello proposto, il fattore RRF interagisce con il ribosoma nel sito A, da dove viene poi spiazzato da EF-G. Questo promuove il rilascio del tRNA, dopodiché RRF, EF-G e l'mRNA vengono rilasciati dal ribosoma. A questa fase di dissociazione contribuisce anche il fattore di inizio IF3, che, come abbiamo visto precedentemente, lega la subunità 30S impedendone la riassociazione con la 50S.

Negli eucarioti la terminazione sembra essere molto simile a quella procariotica, ma vi si trova un solo fattore di rilascio di classe 1 chiamato **eRF1**, che riconosce tutti e tre i codoni di terminazione, e un solo fattore di rilascio di classe 2 chiamato **eRF3**.

15.5 Bilancio energetico

Vari fattori coinvolti nella sintesi proteica richiedono energia sotto forma di GTP. Inoltre, altra energia è necessaria per la formazione del legame peptidico da parte dell'attività peptidil transferasica del ribosoma. Questa energia proviene dalla molecola di ATP che viene idrolizzata dalla amminoacil-tRNA sintetasi durante l'attivazione e il caricamento dall'amminoacido sul tRNA corrispondente. Tale energia viene conservata nel legame tra tRNA e amminoacido per poi essere utilizzata al momento della formazione del legame peptidico. In tutto, quindi, per ogni amminoacido che viene aggiunto alla catena polipeptidica durante la fase di allungamento vengono usate una molecola di ATP, per l'attivazione dell'amminoacido, e due di GTP, una dal fattore EF-Tu/Ts che media l'entrata dell'amminoacil-tRNA nel sito A del ribosoma e l'altra dal fattore EF-G coinvolto nella traslocazione. Si tratta di una notevole quantità di energia ($\Delta G = -22$

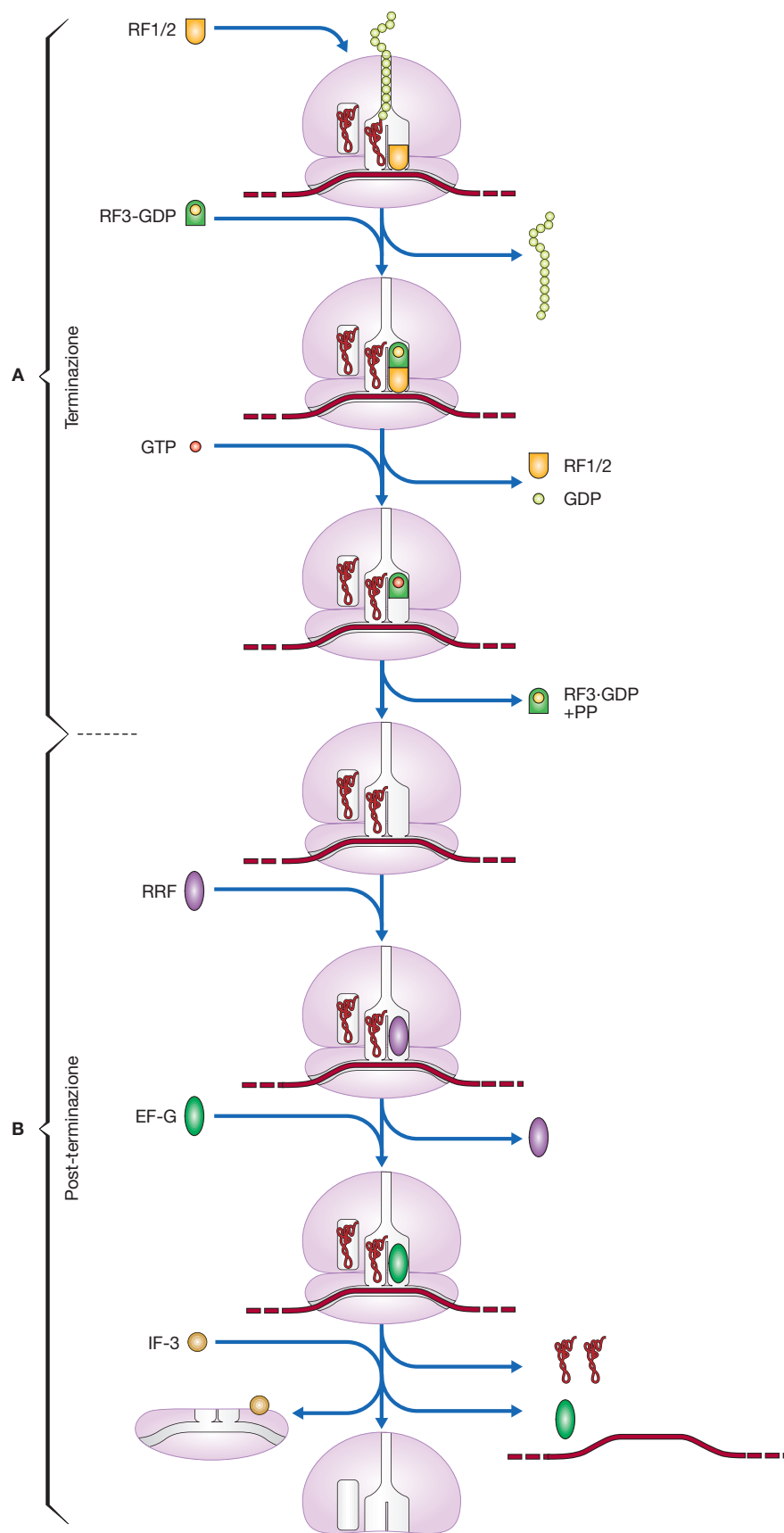


Figura 15.14 Terminazione e post-terminazione della traduzione nei procarioti. (A) Terminazione: RF1 (o RF2) entra nell'emito A della subunità piccola del ribosoma riconoscendo il codone di terminazione e induce il rilascio del polipeptide dal peptidil-tRNA. Successivamente RF3-GDP entra nell'emito A della subunità grande e interagisce con RF1 (o RF2). Lo scambio GDP-GTP provoca il rilascio di RF1 (o RF2) e, successivamente, dello stesso RF3-GDP e pirofosfato. (B) Post-terminazione: il complesso ribosoma/mRNA/tRNA viene disassemblato per effetto del fattore RRF, che interagisce con il sito A del ribosoma, da dove viene poi spazzato da EF-G. Questo porta al rilascio del tRNA e successivamente al disassemblaggio degli altri componenti, a cui contribuisce anche IF3, che, legando la subunità 30S, ne impedisce la riassociazione con la 50S.

FINESTRA 15.2 SISTEMI DI SINTESI PROTEICA *IN VITRO*

I sistemi di sintesi proteica *in vitro* sono un importante strumento di studio per la Biologia Molecolare. Vengono chiamati anche **sistemi di traduzione acellulare** (*cell-free*) perché sono ottenuti rompendo le cellule e prelevando la frazione citoplasmatica che è in grado, se incubata in condizioni opportune, di portare avanti la sintesi proteica per un limitato periodo di tempo. In teoria si possono preparare sistemi di sintesi proteica *in vitro* da qualsiasi tipo di cellule, ma in pratica ne sono stati messi a punto pochi, in genere utilizzando come materiale di partenza cellule impegnate in un'intensa sintesi proteica. I sistemi di sintesi proteica *in vitro* più utilizzati sono l'estratto cellulare di *Escherichia*

coli, il lisato di reticolociti di coniglio e l'estratto di germe di grano. In tutti i casi vengono preparati come estratti grezzi che contengono tutti i componenti molecolari per la traduzione di mRNA esogeni: ribosomi procariotici 70S o eucariotici 80S, tRNA, amminoacil-sintetasi, fattori di inizio, allungamento e terminazione ecc. Per assicurare una traduzione efficiente, l'estratto deve essere addizionato con amminoacidi, sorgenti di energia (ATP, GTP), un sistema per rigenerare l'energia (fosfoenol-piruvato e piruvato chinasi per estratti di *E. coli*; creatin-fosfato e creatina fosfochinasi per i sistemi eucariotici) e altri cofattori (Mg^{2+} , K^+ ecc.).

kcal/mol), molto maggiore di quella che sarebbe richiesta per la formazione di un singolo legame peptidico ($\Delta G = +0,5$ kcal/mol). Già le -8 kcal/mol liberate dall'idrolisi dell'ATP sono più che sufficienti a garantire che l'equilibrio della reazione sia spostato verso la sintesi della proteina piuttosto che verso la degradazione. Perché allora è necessaria la liberazione di tanta altra energia mediante idrolisi di GTP? Il fatto è che i fattori di allungamento EF-Tu/Ts ed EF-G, e anche quelli di inizio IF2 e di terminazione RF3, svolgono il loro ruolo impiegando l'energia di GTP per dirigere complessi cambiamenti conformazionali e movimenti molecolari necessari per la traduzione. In maniera più generica, potremmo dire che solo una piccola parte dell'energia spesa per la sintesi proteica è utilizzata per la formazione dei legami peptidici tra gli amminoacidi. Una gran parte è invece necessaria a far muovere e funzionare il complesso apparato della traduzione che garantisce non solo la semplice unione di amminoacidi ma anche la decodificazione dell'informazione e la sintesi di proteine in cui la sequenza degli amminoacidi è ben determinata. In altre parole, una piccola parte dell'energia va a finire in legami chimici, mentre gran parte va in informazione ed ordine.

15.6 Velocità e accuratezza della sintesi proteica

La velocità di allungamento della catena amminoacidica è, nei procarioti, di circa 15 amminoacidi al secondo ed è circa 10 volte più lenta negli eucarioti. È una velocità molto bassa se la confrontiamo con quella delle comuni reazioni chimiche e anche rispetto alla maggior parte delle reazioni enzimatiche. D'altra parte, appare invece rapida se pensiamo alla complessità della reazione. Infatti, questa comporta, oltre a complessi riarrangiamenti strutturali e movimenti molecolari come quello delle subunità ribosomali, anche la selezione del corretto amminoacil-tRNA tra una collezione di venti tipi diversi, e ciò può essere fatto solo con un susseguirsi di "prova ed errore" in cui i vari aa-tRNA vengono introdotti da EF-Tu·GTP nel sito A del ribosoma per tentare l'appaiamento codone/anticodone. Esiste una relazione inversa tra velocità e accuratezza della sintesi proteica: più velocemente procede la traduzione, tanto meno essa è accurata. La selezione dell'amminoacido corretto viene verificata in due fasi successive, quindi con un processo di "correzione di bozze". Infatti, in una prima fase viene verificato che il complesso ternario, costituito da EF-Tu·GTP e amminoacil tRNA, sia quello corretto corrispondente al codone esposto dall'mRNA nel sito A del ribosoma. Dopo che EF-Tu·GTP ha portato l'amminoacil-tRNA nel sito A, idrolizza il GTP e si dissocia dal ribosoma. Fino a che il GTP non è stato idrolizzato, l'inte-

razione tra amminoacil-tRNA e ribosoma è in genere poco stabile, e soprattutto diventa molto instabile se non si ha un corretto appaiamento codone/anticodone. Così vari complessi ternari vengono provati fino a che un'interazione codone/anticodone corretta rende l'interazione più stabile. Quando poi il GTP viene idrolizzato, l'amminoacil-tRNA viene definitivamente lasciato nel sito A da EF-Tu·GDP che lascia il ribosoma. A questo punto interviene la seconda fase di verifica che caratterizza la correzione di bozze. Infatti, se l'amminoacil-tRNA lasciato da EF-Tu·GDP nel sito A non è quello corretto, esso ha un'alta probabilità di essere rilasciato dal ribosoma, anche questa volta a causa dell'instabilità dell'errato appaiamento codone/anticodone. A questo delicato equilibrio tra velocità e accuratezza della traduzione contribuiscono vari elementi. Uno di questi è, relativamente alla prima fase di verifica, il tasso di idrolisi di GTP da parte di EF-Tu, che non deve essere troppo rapido per dare tempo alla selezione del corretto complesso ternario.

15.7 Traduzione a livello strutturale

Nella descrizione che abbiamo fatto del processo di traduzione sono stati presentati modelli molto semplificati in cui i vari componenti dell'apparato di traduzione sono simbolicamente rappresentati nelle figure come oggetti sferici, ovali ecc. Questo livello di descrizione era l'unico possibile ancora fino a pochi anni fa e comunque è ancora oggi molto utile per spiegare modi e tempi delle interazioni tra i numerosi componenti del sistema, ma sicuramente è una rappresentazione molto poco realistica. D'altra parte la definizione ad alta risoluzione della struttura tridimensionale del ribosoma procariotico, raggiunta nei primi anni 2000, ha permesso di iniziare ad affrontare a livello molecolare più fine lo studio del meccanismo della traduzione, di cui diamo qui solo alcuni esempi. La **figura 15.15** mostra l'interfaccia tra le due subunità 50S e 30S e il posizionamento preciso dei 3 tRNA nei siti A, P ed E. Come si vede, nella figura i 3 siti si trovano a cavallo tra le due subunità, cosicché ciascuna di queste presenta 3 emisiti, A,

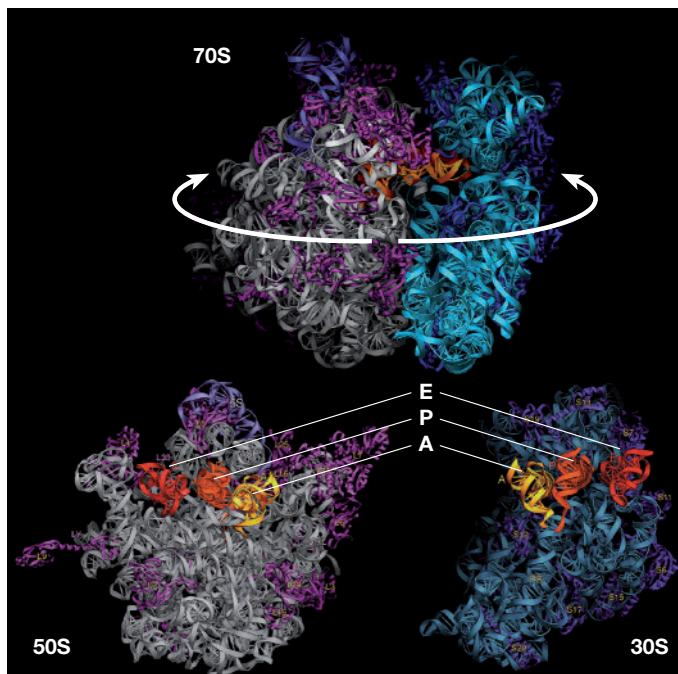


Figura 15.15 Struttura ad alta risoluzione del ribosoma e dei siti A, P ed E per i tRNA. Un ribosoma 70S viene aperto mettendo in mostra l'interfaccia tra le due subunità (le due facce che sono in contatto). Si possono osservare gli emisiti A, P ed E, che nelle due subunità alloggiavano rispettivamente l'amminoacil-tRNA, il peptidil-tRNA e il tRNA scarico in uscita. (Fonte: Center for Molecular Biology of RNA, University of California, Santa Cruz, http://rna.ucsc.edu/rnacenter/noller_lab.html.)

FINESTRA 15.3

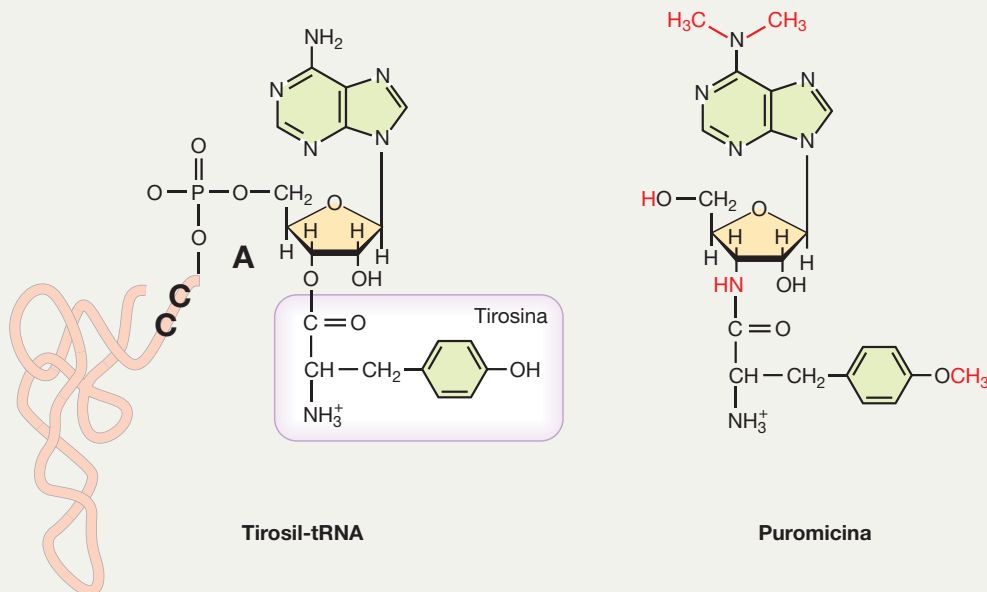
INIBITORI DELLA SINTESI PROTEICA E ANTIBIOTICI

Lo studio del meccanismo della traduzione, come d'altra parte anche della replicazione del DNA e della trascrizione, si è avvantaggiato dell'uso di inibitori specifici capaci d'interferire in vario modo con i diversi passaggi del processo. Infatti, la possibilità di inibire o alterare una reazione permette di studiarne il meccanismo. Tra questi hanno un ruolo preponderante gli antibiotici, che hanno anche un grande interesse applicativo per l'industria farmaceutica. Molti degli antibiotici maggiormente usati in terapia sono inibitori di qualche componente del processo di sintesi proteica e, in particolare, è spesso coinvolta un'interazione dell'antibiotico con il ribosoma. I più interessanti dal punto di vista terapeutico sono quelli che inibiscono la sintesi proteica procariotica ma non quella eucariotica, permettendo quindi di fermare la crescita batterica senza effetti negativi sulle cellule dell'organismo eucariotico infettato. Alcuni altri antibiotici inibiscono la sintesi proteica sia procariotica che eucariotica, e altri ancora sono specifici per gli eucarioti. Anche questi possono trovare alcune applicazioni per inibire la crescita di organismi eucariotici inferiori, per esempio funghi, mediante farmaci "per uso esterno". Tutti questi tipi di antibiotici sono stati e sono tutt'ora utili per elucidare i meccanismi molecolari della sintesi proteica. Vediamone alcuni esempi. La **puromicina** è una molecola la cui struttura chimica è simile all'estremità del braccio accettore di un amminoacil-tRNA, in particolare di un tirosil-RNA. Questa somiglianza permette alla puromicina di entrare nell'emisita A presente nella subunità grande del ribosoma, dove si trova il centro della peptidil transferasi. Questa catalizza il trasferimento della catena polipeptidica nascente sulla puromicina, che però, mancando della restante parte strutturale corrispondente all'intero tRNA, non mantiene il contatto con il ribosoma e viene rilasciata, causando così una terminazione prematura

dell'allungamento e il rilascio di peptidi incompleti dal ribosoma. Come ci si sarebbe potuti aspettare, la puromicina inibisce allo stesso modo la sintesi proteica procariotica ed eucariotica.

L'antibiotico steroideo **acido fusidico** agisce solo sui procarioti Gram-positivi, bloccandone i ribosomi subito dopo la traslocazione. In presenza di acido fusidico la traslocazione può avvenire regolarmente: infatti EF-G si lega al ribosoma, il GTP viene idrolizzato e il ribosoma si sposta di tre nucleotidi sull'mRNA. A questo punto, però, l'acido fusidico stabilizza il complesso ribosoma/EF-G/GDP impedendo che EF-G e GDP vengano rilasciati. Pertanto, poiché il ribosoma non può legare un nuovo amminoacil-tRNA, non possono essere aggiunti nuovi amminoacidi alla catena peptidica. La **streptomicina** interagisce con la subunità 30S del ribosoma procariotico, in particolare con la r-proteina S12. Ad alte concentrazioni questo antibiotico blocca la sintesi proteica inibendo la transizione da complesso di inizio a fase di allungamento. A concentrazioni limitate, in cui la traduzione può ancora essere portata avanti, il legame di questo antibiotico con la subunità piccola del ribosoma determina un'alterazione strutturale del centro di decodificazione del ribosoma che causa un aumento degli errori di lettura del codice e quindi l'inserimento di amminoacidi sbagliati nel peptide sintetizzato.

Menzioniamo brevemente alcuni dei numerosi altri inibitori della sintesi proteica. Il **cloramfenicolo** lega la subunità ribosomiale 50S procariotica bloccando l'attività peptidil transferasica e quindi il trasferimento del peptide nascente all'amminoacil-tRNA sul sito A, mentre la **cicloesimide** inibisce l'attività della peptidil transferasi negli eucarioti. La **tetraciclina** inibisce nei procarioti il legame dell'amminoacil-tRNA al sito A.



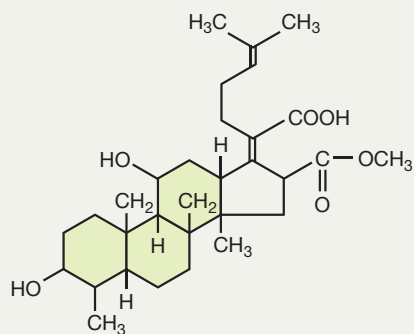
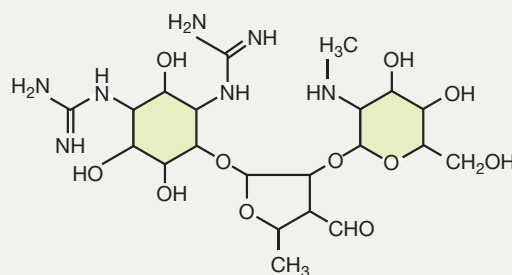
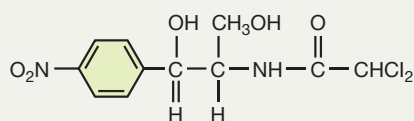
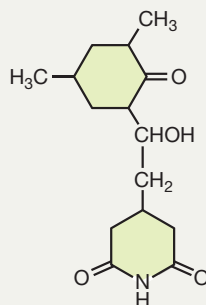
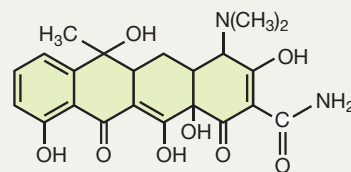
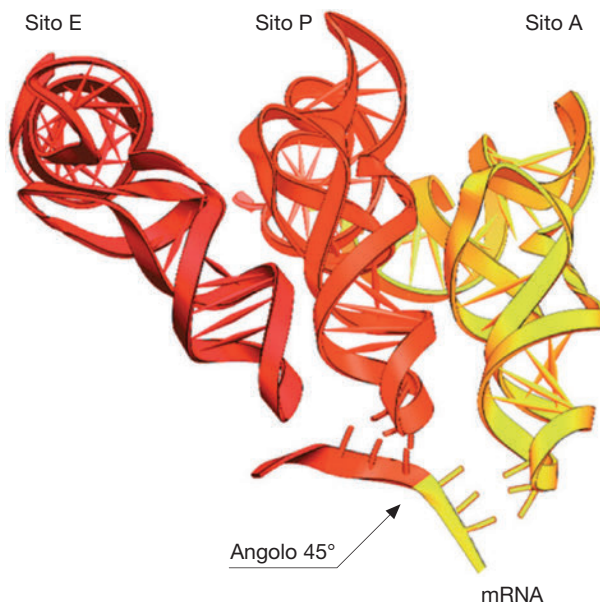
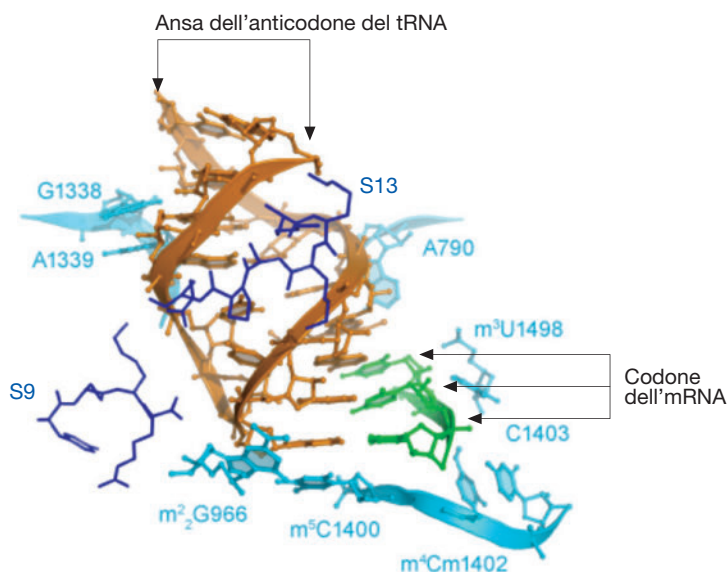
**Acido fusidico****Streptomicina****Cloramfenicolo****Cicloesimide****Tetraciclina**

Figura 15.16 Posizionamento relativo dei tRNA che si trovano nei siti A, P ed E del ribosoma. L'mRNA fa una piega di 45° che consente ai tRNA nei siti A e P di appaiarsi a codoni adiacenti sull'mRNA. (Fonte: Center for Molecular Biology of RNA, University of California, Santa Cruz, http://rna.ucsc.edu/rnacenter/noller_lab.html.)

P ed E. La **figura 15.16** mostra invece le relazioni all'interno della struttura del ribosoma (che qui è omessa) tra gli anticodoni dei tRNA e i codoni sull'mRNA. Le molecole di tRNA, che sono abbastanza ingombranti, devono posizionarsi nei siti P e A in maniera tale da poter appaiare contemporaneamente i loro anticodoni con due codoni adiacenti sull'mRNA. Perché ciò sia possibile, l'mRNA che si trova nel canale dell'mRNA si piega con un angolo di circa 45° nella regione del centro di decodificazione. La **figura 15.17** mostra nel dettaglio l'interazione tra l'ansa dell'anticodone del tRNA^{Phe} e il codone dell'mRNA nel sito P della subunità ribosomale 30S.

Figura 15.17 Particolare della struttura del ribosoma di *Thermus thermophilus* alla risoluzione di 3,7 Å. Interazione, nel sito P della subunità ribosomale 30S, tra l'ansa dell'anticodone del tRNA (marrone), il codone dell'mRNA (verde), tratti delle r-proteine S9 e S13 (blu) e alcune basi della sequenza dell'rRNA16S (celeste). [Fonte: Korostelev, A. *et al.* (2006), *Cell*, 126, pp. 1065-1067.]



Lecture di approfondimento

- Jackson, R.J. (2005), «Alternative mechanisms of initiating translation of mammalian mRNAs», *Biochemical Society Transactions*, **33**, pp. 1231-1241.
- Kapp, L.D. e Lorsch, J.R. (2004), «The molecular mechanics of eukaryotic translation», *Annual Review of Biochemistry*, **73**, pp. 657-704.
- Laursen, B.S., Sørensen, H.P., Mortensen, K.K. e Sperling-Petersen, H.U. (2005), «Initiation of protein synthesis in bacteria», *Microbiology and Molecular Biology Reviews*, **69**, pp. 101-123.
- Schmeing, T.M. e Ramakrishnan, V. (2009), «What recent ribosome structures have revealed about the mechanism of translation», *Nature*, **461**, pp. 1234-1242.

16

Regolazione della traduzione

16.1 Regolazione generale della traduzione

16.2 Regolazioni traduzionali di geni specifici

16.3 Regolazione della stabilità e degradazione degli mRNA

16.4 Trasporto e localizzazione degli mRNA

16.5 Granuli citoplasmatici di RNA

Come tutti i processi implicati nel trasferimento dell'informazione genica, anche la traduzione è un passaggio dove operano vari meccanismi regolativi. Esistono innanzitutto meccanismi di regolazione dell'efficienza della traduzione degli mRNA che possono riguardare tutti gli mRNA della cellula o solo alcuni di essi. Altri meccanismi di regolazione riguardano la stabilità degli mRNA. Infatti, a parità di espressione genica a livello trascrizionale, un mRNA stabile produrrà più prodotto proteico rispetto a un mRNA che ha una vita media breve, oppure la sintesi di una proteina può essere rapidamente interrotta da un meccanismo che porta a rapida degradazione il corrispondente mRNA. Un altro tipo di regolazione riguarda la localizzazione di alcuni mRNA che vanno a posizionarsi in particolari regioni del citoplasma dove vengono tradotti, contribuendo così all'asimmetria strutturale della cellula. Anche se sicuramente questi vari fenomeni sono interconnessi, non abbiamo ancora una visione generale della situazione e possiamo qui solo dare delle descrizioni parziali.

16.1 Regolazione generale della traduzione

“Risposta stringente” nei procarioti

Quando una cellula procariotica si viene a trovare in condizioni di crescita sfavorevoli e la concentrazione di amminoacidi nel mezzo non è sufficiente a sostenere un normale tasso di sintesi proteica, essa attiva un meccanismo di difesa chiamato **risposta stringente** che porta alla riduzione di varie funzioni in modo da economizzare risorse e sopravvivere fino a tempi migliori. La risposta stringente comporta tra l'altro una drastica riduzione della sintesi di tRNA e rRNA e, quindi, una diminuita produzione dell'apparato di traduzione, che scende a meno del 10% del normale. Si ha anche una riduzione della sintesi di mRNA e altre alterazioni metaboliche, tra cui la riduzione della sintesi di carboidrati, lipidi ecc. È stato trovato che la risposta stringente viene innescata dall'ingresso, nel sito A del ribosoma, di un tRNA scarico, ma corretto rispetto al codone esposto dall'mRNA e questo avviene quando, per carenza di amminoacidi, la maggior parte dei tRNA sono scarichi. Al ribosoma, o meglio

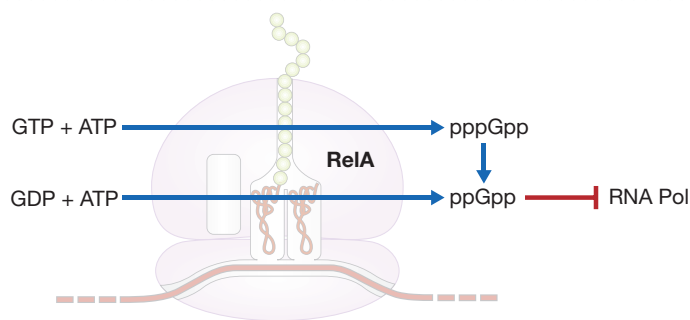


Figura 16.1 Regolazione generale della traduzione nei procarioti: la “risposta stringente”. In condizioni di crescita sfavorevoli per carenza di amminoacidi, l'ingresso di un tRNA scarico nel sito A del ribosoma innesca la risposta stringente da parte del fattore RelA. RelA, che è associato ai ribosomi, “sente” la presenza del

tRNA scarico nel sito A e catalizza, attraverso le due vie indicate, la sintesi di ppGpp. Questo va a modificare l'attività di varie proteine bersaglio e, in particolare, interagisce con l'RNA polimerasi riducendo il tasso generale di trascrizione e, quindi, anche la produzione di nuovi ribosomi.

solo ad alcuni ribosomi, è associata la proteina RelA, chiamata anche **fattore di stringenza**. Questa, in risposta alla presenza del tRNA scarico nel sito A, catalizza la reazione mostrata in **figura 16.1**, in cui l'ATP dona un gruppo pirofosfato alla posizione 3' del GDP o del GTP, portando alla formazione di due nucleotidi anomali, ppGpp (guanosina tetrafosfato) e pppGpp (guanosina pentafrasfato). Il primo ha due gruppi difosfato, uno nella posizione 5' e l'altro in 3'; il secondo ha un gruppo trifosfato in 5' e un difosfato in 3'. pppGpp viene poi convertito in ppGpp, che è il reale effettore della risposta stringente in quanto si lega a varie proteine bersaglio alterandone l'attività. L'interazione di ppGpp con l'RNA polimerasi causa una riduzione del tasso di trascrizione in generale e in modo particolare della trascrizione dei geni per l'rRNA e, quindi, della produzione di ribosomi. In questo modo la cellula regola la produzione di apparato traduzionale in relazione alle condizioni di crescita più o meno favorevoli.

Controllo generale della traduzione negli eucarioti

Anche le cellule eucariotiche devono poter regolare l'efficienza della traduzione in relazione alla disponibilità di nutrienti quali gli amminoacidi. Inoltre, soprattutto le cellule degli eucarioti superiori devono rispondere ai vari segnali di crescita cellulare, o di blocco della crescita, che arrivano dall'esterno sotto forma di ormoni, fattori di crescita ecc. Questi interagiscono con recettori specifici della cellula e vanno ad attivare vie di trasduzione del segnale, alcune delle quali raggiungono il nucleo, dove attivano la trascrizione di specifici geni bersaglio, mentre altre influenzano, nel citoplasma, l'efficienza della sintesi proteica e quindi la crescita cellulare. Non ci addentriamo qui nei processi di trasduzione del segnale per i quali si rinvia ad altre fonti. Ci basti dire che questi meccanismi di regolazione globale della traduzione si basano in gran parte sulle modificazioni post-traduzionali, soprattutto fosforilazioni (*vedi* cap. 20), di fattori di inizio o di proteine che legano i fattori di inizio. Per esempio, come illustrato in **figura 16.2**, la fosforilazione del fattore eIF2 α , operata da specifiche chinasi quando le cellule sono deprivate di amminoacidi o soggette a vari tipi di stress (shock termico, ipossia, raggi ultravioletti ecc.), risulta in un blocco della formazione del complesso ternario eIF2-GTP/aa-tRNA e quindi in un'inibizione di tutta la sintesi proteica. Anche il fattore di inizio eIF4E, che si lega al 5'cap degli mRNA ed è implicato nell'inizio della traduzione cap-dipendente, è soggetto a fosforilazione. In questo caso la fosforilazione avviene solo in presenza di fattori di crescita e risulta in una stimolazione, anziché in una repressione, della tradizio-

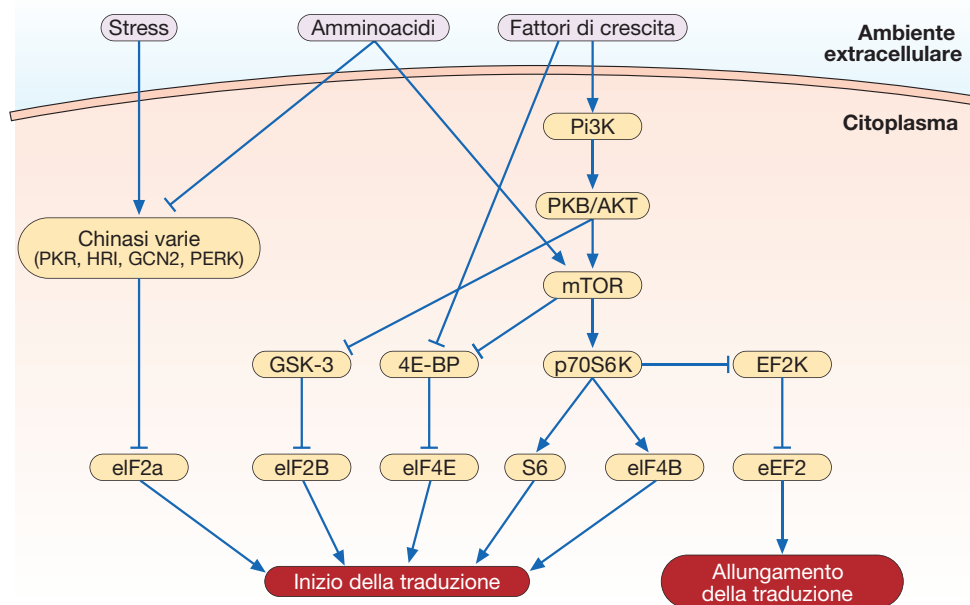


Figura 16.2 Regolazione generale della traduzione negli eucarioti. L'efficienza generale della traduzione è influenzata, a livello di inizio e di allungamento, da segnali extracellulari (nutrienti, fattori di crescita, particolari condizioni ambientali ecc.). Le vie di trasduzione del segnale coinvolte sono in parte sovrapposte e intersecate con quelle responsabili, nel nucleo, della regolazione della trascrizione e replicazione del DNA.

Nel citoplasma queste vie di segnalazione, che consistono in cascate di modificazioni post-traduzionali (perlopiù fosforilazioni), influenzano l'efficienza della traduzione mediante attivazione o inibizione di fattori di inizio e di allungamento. Le frecce con estremità a punta indicano "attivazione", le frecce con estremità piatta indicano "inibizione".

ne. Infatti, eIF4E è inibito dal legame di altri fattori proteici, chiamati **4E-BP** (4E Binding Protein), che in presenza di fattori di crescita vengono fosforilati e si staccano da eIF4E, con conseguente stimolazione della traduzione. Questi pochi esempi danno un'idea di come segnali esterni o intracellulari possano regolare l'efficienza generale della sintesi proteica. In realtà questi meccanismi di regolazione traduzionale generale sono molto più complessi e per molti aspetti ancora poco chiari.

16.2 Regolazioni traduzionali di geni specifici

Benché il controllo dell'espressione genica avvenga principalmente a livello della trascrizione, cioè regolando l'efficienza della sintesi degli mRNA, essa può essere controllata anche a livello traduzionale, cioè regolando l'efficienza di utilizzazione di specifici mRNA da parte dei ribosomi. L'importanza di questo livello di controllo sta diventando sempre più evidente sia nei procarioti che negli eucarioti e lo si trova implicato in varie situazioni fisiologiche, tra cui lo sviluppo embrionale e lo sviluppo e funzione del sistema nervoso. Ciò che rende la regolazione traduzionale particolarmente efficace è il fatto che si tratta di una regolazione molto rapida e reversibile: pochi minuti sono sufficienti per bloccare la sintesi di una proteina il cui mRNA può poi, in pochi minuti, essere riattivato. I meccanismi molecolari alla base della regolazione traduzionale sono molto vari e si continuano a trovare nuove interessanti situazioni. Consideriamo innanzitutto casi relativamente semplici in cui la regolazione specifica della traduzione è dovuta a fattori proteici che si legano agli mRNA e ne impediscono il legame con i ribosomi, funzionando così da "repressori traduzionali", analogamente ai classici repressori della trascrizione che legandosi al DNA a monte dei geni ne controllano l'efficienza di trascrizione.

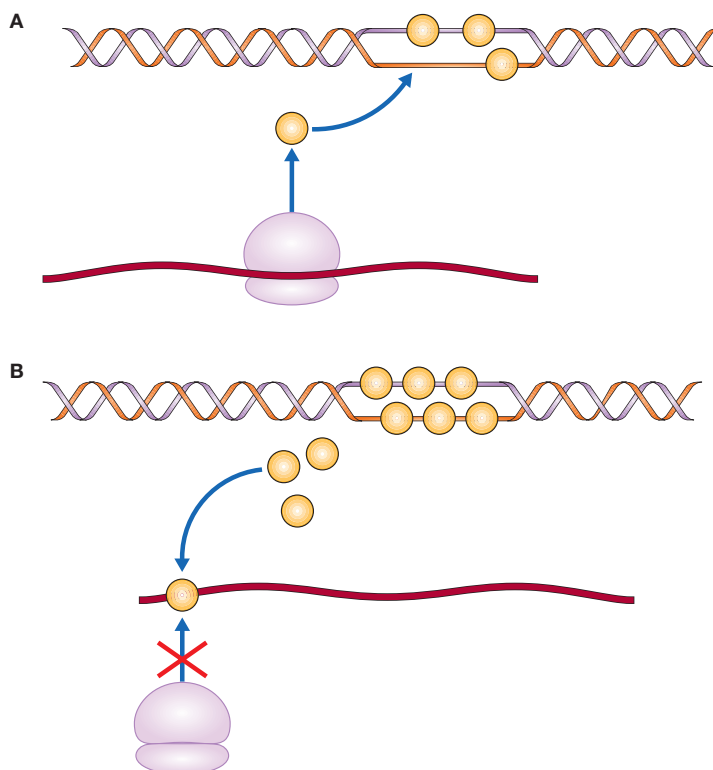
In alcuni casi, soprattutto nei procarioti, si tratta di controlli autogeni (feedback) in cui il prodotto finale, se in eccesso, va a bloccare la propria sintesi a livello di traduzione; in altri casi si tratta invece di controlli da parte di fattori prodotti da geni regolatori che agiscono in *trans* sulla traduzione di altri mRNA. Descriveremo qui alcuni esempi di regolazione traduzionale, autogeni e non, nei procarioti e negli eucarioti.

Controlli traduzionali autogeni

Il gene 32 del fago T4 codifica una proteina, chiamata **p32**, capace di legare il DNA a singolo filamento e che interviene in vari processi dell'infezione fagica, quali la replicazione, la ricombinazione e la riparazione del DNA. Oltre ad avere un'alta affinità per il DNA a singolo filamento, p32 mostra una certa affinità anche per l'RNA a singolo filamento e in particolare per una regione della 5'UTR del suo stesso mRNA, e quando vi è legata ne blocca la traduzione. Un sistema così congegnato risulta in un semplice, ma molto efficiente, controllo autogeno (fig. 16.3). Infatti, p32 sarà sempre prodotta nella quantità necessaria a legare il DNA a singolo filamento perché, non appena ne viene prodotta in eccesso, andrà a legare anche il proprio mRNA bloccandone l'ulteriore traduzione. Naturalmente, affinché il sistema funzioni, è necessario che l'affinità di p32 per il proprio mRNA sia minore di quella per il DNA a singolo filamento, in modo che si abbia inibizione della traduzione solo quando il DNA a singolo filamento sia saturato da p32.

Un caso un po' più complicato di controllo autogeno a livello traduzionale lo troviamo per i geni che codificano le proteine ribosomali di *Escherichia coli*. Il ribosoma di *E. coli* contiene 55 r-proteine diverse nelle sue subunità 30S e 50S. I geni che codificano queste proteine sono raggruppati in alcuni operoni policistronici che contengono da 2 a oltre 10 geni per diverse r-proteine (in realtà in questi

Figura 16.3 Regolazione traduzionale autogena della sintesi della proteina p32 del fago T4. (A) Il ruolo della proteina p32, codificata dal gene 32 del fago T4, è quello di legare regioni a singolo filamento del DNA del fago stesso. (B) Se p32 viene prodotta in eccesso, dopo avere saturato le regioni a singolo filamento va a legare anche la 5'UTR del proprio mRNA, bloccandone così la traduzione.



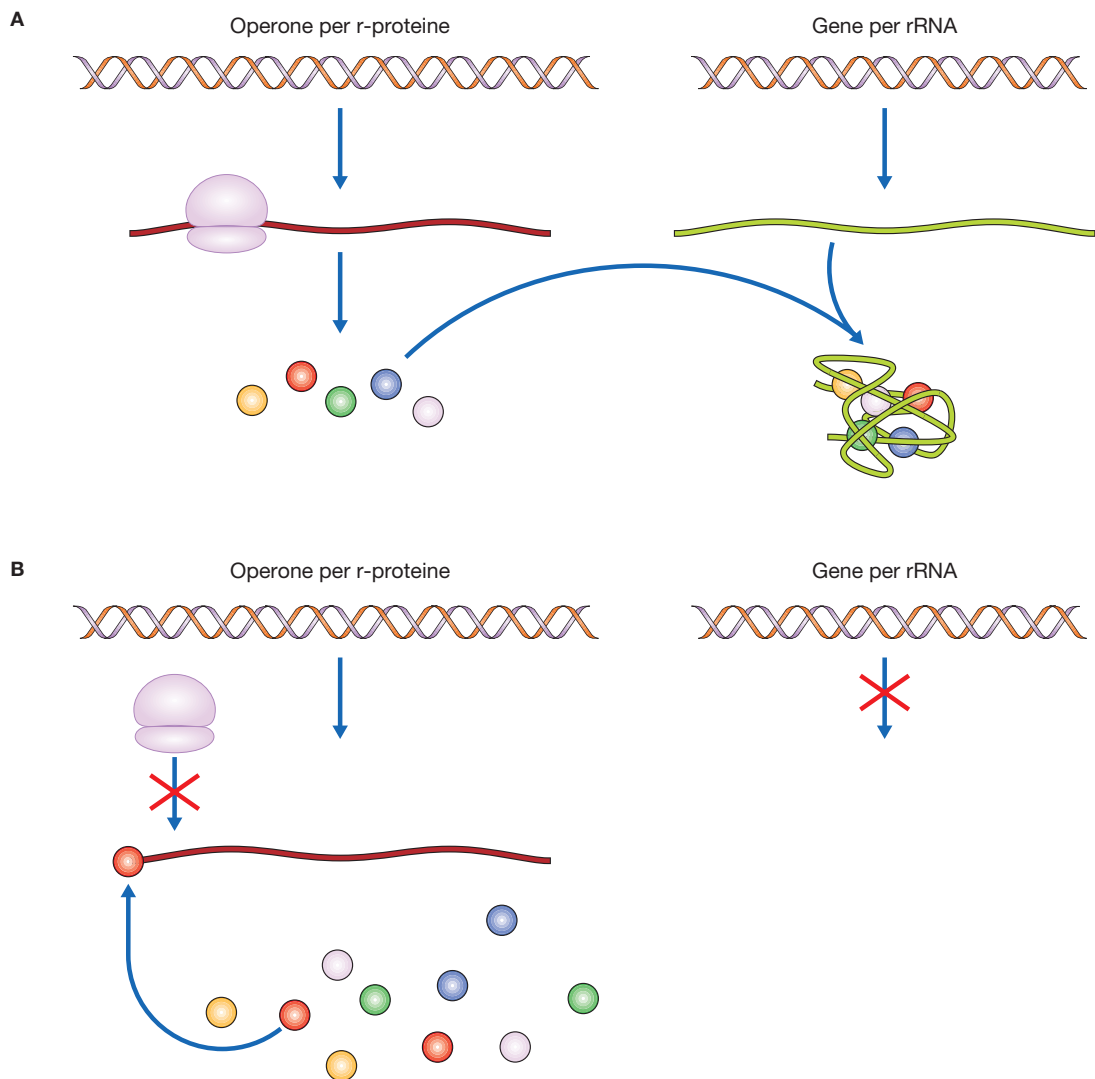


Figura 16.4 Regolazione traduzionale autogena della sintesi delle r-proteine di *Escherichia coli*. I geni che codificano per queste proteine sono raggruppati in alcuni operoni policistronici, dei quali ne è qui rappresentato uno codificante per 5 r-proteine. (A) Finché la quantità di r-proteine sintetizzate è bilanciata con la quantità di rRNA prodotta, la traduzione continua. (B) Quando la quantità di r-

proteine è in eccesso rispetto all'rRNA con cui devono assemblarsi, magari perché la trascrizione dell'rRNA è rallentata o bloccata, una delle r-proteine in eccesso va a legarsi all'mRNA bloccandone l'ulteriore traduzione, che può eventualmente riprendere quando la situazione si sia riequilibrata.

operoni si trovano anche i geni per alcune altre proteine coinvolte nella sintesi proteica). Ciascuno di questi operoni è regolato traduzionalmente mediante un meccanismo autogeno da una delle r-proteine codificata dall'operone stesso. La **figura 16.4** mostra schematicamente uno di questi operoni che codifica per 5 diverse r-proteine. Queste, una volta sintetizzate, vanno a legarsi, insieme a tutte le altre r-proteine, all'rRNA nascente per assemblare nuove subunità ribosomiali (**fig. 16.4A**). Se però sono in eccesso rispetto all'rRNA con cui assemblarsi, magari perché per qualche motivo è stata rallentata o bloccata l'attività del gene per l'rRNA, una di esse va a legarsi alla regione 5'UTR del loro stesso mRNA bloccandone l'ulteriore traduzione (**fig. 16.4B**). Quindi il modello è sostanzialmente uguale a quello sopra descritto per la proteina p32 del fago T4, con la differenza che una proteina regola la sintesi di se stessa e delle altre codificate dallo stesso operone.

Un esempio di regolazione traduzionale autogena negli eucarioti è quello della **PABP** [*Poli(A) Binding Protein* = proteina che lega il poli(A)]. Come abbiamo già ricordato questa proteina lega la coda di poli(A) presente all'estremità 3' di tutti gli mRNA eucariotici e ha un ruolo nella stabilizzazione dell'mRNA. Inoltre, poiché interagisce anche con il fattore eIF4G (*vedi* fig. 15.9), porta a contatto le due estremità dell'mRNA, che assume così una conformazione circolare e può svolgere un ruolo anche nella fase di inizio della traduzione. È stato osservato che l'mRNA che codifica la PABP contiene, oltre alla coda di poli(A), anche una regione ricca di tratti di poli(A) nella 5'UTR. Se la PABP viene sintetizzata in eccesso rispetto alla quantità necessaria, va a legare anche le regioni ricche in A della 5'UTR del proprio mRNA, bloccando così il processo di scansione da parte della subunità piccola del ribosoma e, quindi, inibendo la sintesi della PABP stessa.

Possiamo in genere osservare che la regolazione traduzionale autogena riguarda spesso proteine che vanno a far parte di complessi macromolecolari. Infatti, in questi casi la regolazione autogena garantisce una produzione bilanciata dei vari componenti del complesso, ciascuno dei quali, quando in eccesso, va a rallentare la propria sintesi.

Controlli traduzionali non autogeni

Come abbiamo detto precedentemente, non tutti i controlli traduzionali sono autogeni. Spesso la regolazione della traduzione di specifici mRNA è dovuta all'interazione di una particolare sequenza dell'mRNA con uno specifico repressore traduzionale, diverso dalla proteina codificata dall'mRNA stesso. Un esempio ben studiato di controllo traduzionale non autogeno negli eucarioti è quello della sintesi della ferritina. Il ferro è un elemento importante in vari processi cellulari ma al tempo stesso può essere tossico se presente in eccesso libero nella cellula. Pertanto, la cellula è dotata sia di proteine che mediano l'importazione di ferro all'interno della cellula stessa sia di proteine che sequestrano il ferro entrato, ma non utilizzato. A quest'ultimo scopo è deputata la **ferritina**, una proteina multimerica sferica, costituita da 24 monomeri di ferritina, che protegge la cellula sequestrando il ferro al suo interno. Poiché la ferritina stessa può essere dannosa se prodotta in eccesso, è necessario un meccanismo di regolazione che ne controlli costantemente la sintesi a seconda delle necessità. È stato dimostrato che questa regolazione avviene a livello di traduzione. Infatti, l'mRNA per la ferritina è sempre presente in eccesso nelle cellule, ma la sua traduzione è più o meno efficiente o repressa a seconda della quantità di ferro disponibile.

Qual è il meccanismo molecolare alla base di questa regolazione? Come mostrato in **figura 16.5**, la 5'UTR dell'mRNA per la ferritina contiene, adiacente al 5'cap, una struttura secondaria a forcina, chiamata IRE (*Iron Responsive Element* = elemento di risposta al ferro); questa viene riconosciuta e legata dalla proteina regolatrice IRP (*Iron Regulated Protein* = proteina regolata dal ferro). In carenza di ferro IRP lega la sequenza IRE e impedisce il legame della subunità ribosomale 40S al 5'cap, risultando in un'inibizione della traduzione dell'mRNA e, quindi, in un blocco della sintesi della ferritina. Se il livello di ferro intracellulare aumenta, questo va a legare IRP e ne induce un cambiamento conformazionale che la rende incapace di legare IRE. Questo cambiamento permette il legame della subunità 40S al 5' dell'mRNA e si ha quindi sintesi di ferritina. Si noti l'analogia tra questa regolazione a livello traduzionale e quella dell'operone lattosio a livello trascrizionale: infatti, la modulazione della funzione regolativa di IRP (repressore traduzionale) da parte del ferro è del tutto analoga alla modulazione della funzione del repressore (trascrizionale) dell'operone *lac* da parte del lattosio. Dato questo meccanismo di base, esistono numerose variazioni sul tema. Le sequenze

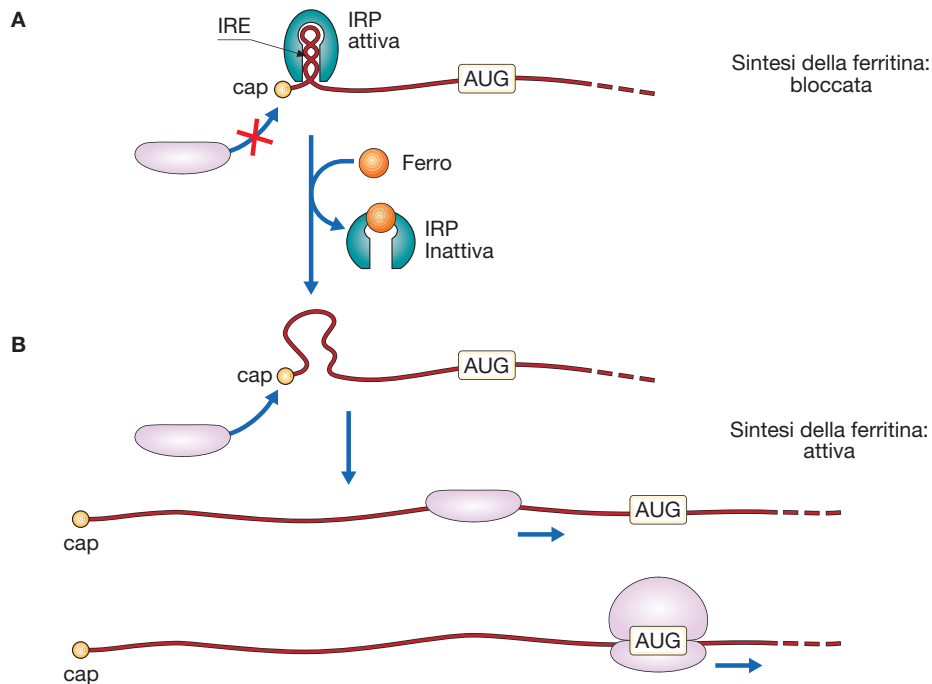


Figura 16.5 Regolazione traduzionale della sintesi della ferritina. (A) In assenza di ferro il repressore traduzionale IRP lega la struttura IRE della 5'UTR dell'mRNA per la ferritina e impedisce il legame della subunità ribosomiale 40S al 5'cap dell'mRNA, bloc-

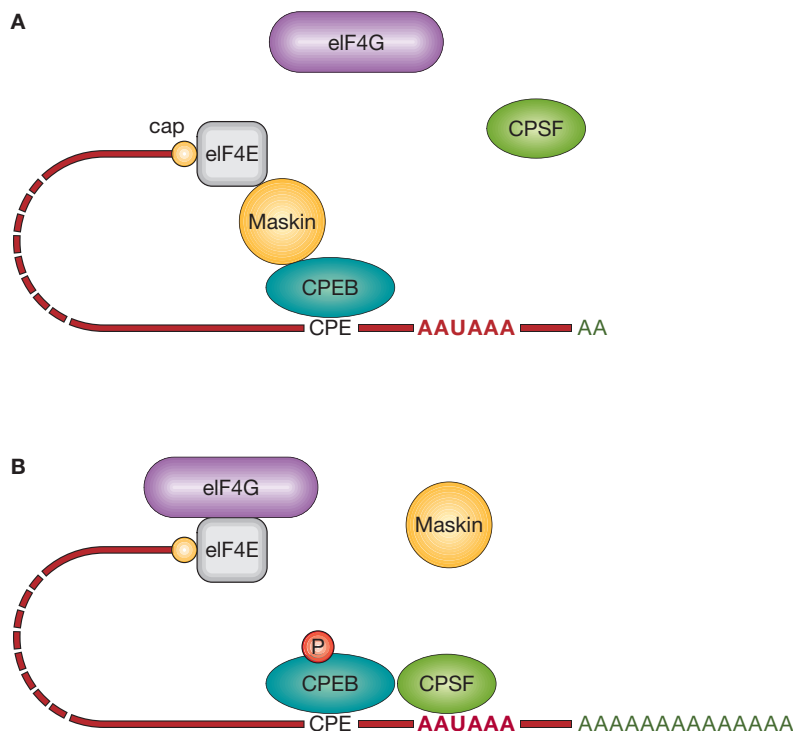
cando così la traduzione. (B) In presenza di ferro, questo lega e inattiva l'IRP, che si dissocia da IRE sbloccando così l'inizio della traduzione.

bersaglio della regolazione non sono sempre localizzate nella 5'UTR, ma spesso anche nella 3'UTR, come d'altra parte ci si può aspettare se consideriamo la "circularizzazione" degli mRNA dovuta all'interazione PABP/eIF4G. Inoltre, spesso non è coinvolto un singolo repressore proteico, ma possono esservi più proteine che interagiscono con l'mRNA e/o tra di loro, alcune con effetto positivo, altre con effetto negativo sulla traduzione.

Un caso interessante è stato studiato nell'oocita di *Xenopus*, dove la regolazione traduzionale passa attraverso il controllo della lunghezza della coda di poli(A) degli mRNA. Infatti, durante l'oogenesi l'oocita accumula molti mRNA (mRNA materni) che non vengono immediatamente tradotti ma accumulati nel citoplasma per essere poi attivati durante i vari stadi dello sviluppo embrionale. Il meccanismo di questo controllo è stato approfondito nel caso dell'mRNA per la ciclina B. La repressione traduzionale di questo mRNA è dovuta a un accorciamento della coda di poli(A), che viene ridotta a circa 20 nt. Questo accorciamento è mediato dalla proteina CPEB, che, come illustrato in **figura 16.6**, lega un elemento di sequenza chiamato CPE (UUUUUAU) che si trova nella 3'UTR dell'mRNA, poco a monte del segnale di poliadenilazione AAUAAA. Come illustrato in **figura 16.6A**, CPEB legato al CPE interagisce con un'altra proteina, la Maskin, che a sua volta interagisce con il fattore di inizio eIF4E. Questa interazione impedisce quella tra eIF4E ed eIF4G, risultando in un blocco dell'inizio della traduzione. In seguito all'attivazione degli oociti, CPEB viene fosforilato da una chinasi (Eg2) e così stimola la poliadenilazione dell'mRNA e la dissociazione di Maskin da CPEB e da eIF4E (**fig. 16.6B**). A questo punto quest'ultimo è libero di interagire con eIF4G e la traduzione può avere inizio.

Gli esempi fin qui riportati, e gli altri che presenteremo più avanti (micro-RNA), si basano su un meccanismo generale abbastanza lineare: un repressore

Figura 16.6 Modello per il controllo dell'inizio della traduzione dell'mRNA per la ciclina B da parte di Maskin. (A) In oociti quiescenti di *Xenopus* la sequenza CPE che si trova nella 3'UTR dell'mRNA per la ciclina B lega la proteina CPEB. Questa induce un drastico accorciamento della coda di poli(A) e interagisce con la proteina Maskin, che, legandosi a sua volta a eIF4E, ne impedisce l'interazione con eIF4G, con conseguente blocco dell'inizio della traduzione. (B) In seguito all'attivazione degli oociti CPEB viene fosforilato, il che permette il reclutamento di CPSF e quindi l'allungamento della coda di poli(A); la fosforilazione di CPEB provoca anche la dissociazione di Maskin da eIF4E, che, essendo ora in grado di interagire con eIF4G, ripristina l'inizio della traduzione. (Per semplicità è stata omessa l'interazione tra la coda di poli(A) e il fattore eIF4G.)



traduzionale lega specificamente un mRNA bersaglio e ne impedisce la traduzione. Esistono però casi di controllo traduzionale basati su meccanismi molto diversi. Prendiamo in considerazione l'esempio ben studiato del gene *GCN4* di lievito. Questo codifica per un fattore trascrizionale che regola positivamente l'espressione di vari geni necessari per la sintesi degli amminoacidi. È stato osservato che la sintesi di *GCN4* stessa è regolata a livello traduzionale dalla presenza di amminoacidi. Infatti, quando la cellula soffre una carenza di amminoacidi, l'mRNA per *GCN4* viene normalmente tradotto portando così all'espressione dei geni implicati nella produzione di nuovi amminoacidi e nella loro utilizzazione.

Quando invece i livelli di amminoacidi nella cellula sono alti, la traduzione dell'mRNA per *GCN4* risulta inibita. La **figura 16.7** illustra il meccanismo alla base di questa regolazione. Analizzando la sequenza nucleotidica dell'mRNA di *GCN4* è stato trovato che, a monte della sequenza codificante per *GCN4*, sono presenti quattro brevissime ORF chiamate uORF1 → uORF4 (*upstream ORF* = ORF a monte) (**fig. 16.7A e B**). Ciascuna di queste è costituita da un codone di inizio AUG seguito da uno o due codoni per amminoacidi e poi da un codone di terminazione. È da notare che questa è una situazione abbastanza anomala per una cellula eucariotica, dove il meccanismo di inizio della traduzione, mediante scansione a partire dal 5' cap, fa sì che in genere il primo AUG sia quello di inizio della sequenza codificante. Comunque, nel caso dell'mRNA di *GCN4* la subunità 40S del ribosoma che procede per scansione partendo dal 5' cap riconosce efficientemente l'AUG della prima uORF, cosicché si ha un evento di inizio con reclutamento della subunità 60S ma, essendoci solo uno o due codoni, la traduzione arriva subito a terminazione.

Come mostrato in **figura 16.7C**, al momento del rilascio, la subunità piccola ha il 50% delle probabilità di restare associata all'mRNA e quindi, previa interazione con il complesso ternario Met-tRNA^{Met}/eIF2/GTP, la possibilità di riconoscere l'AUG dell'uORF subito a valle e di reiniziare.

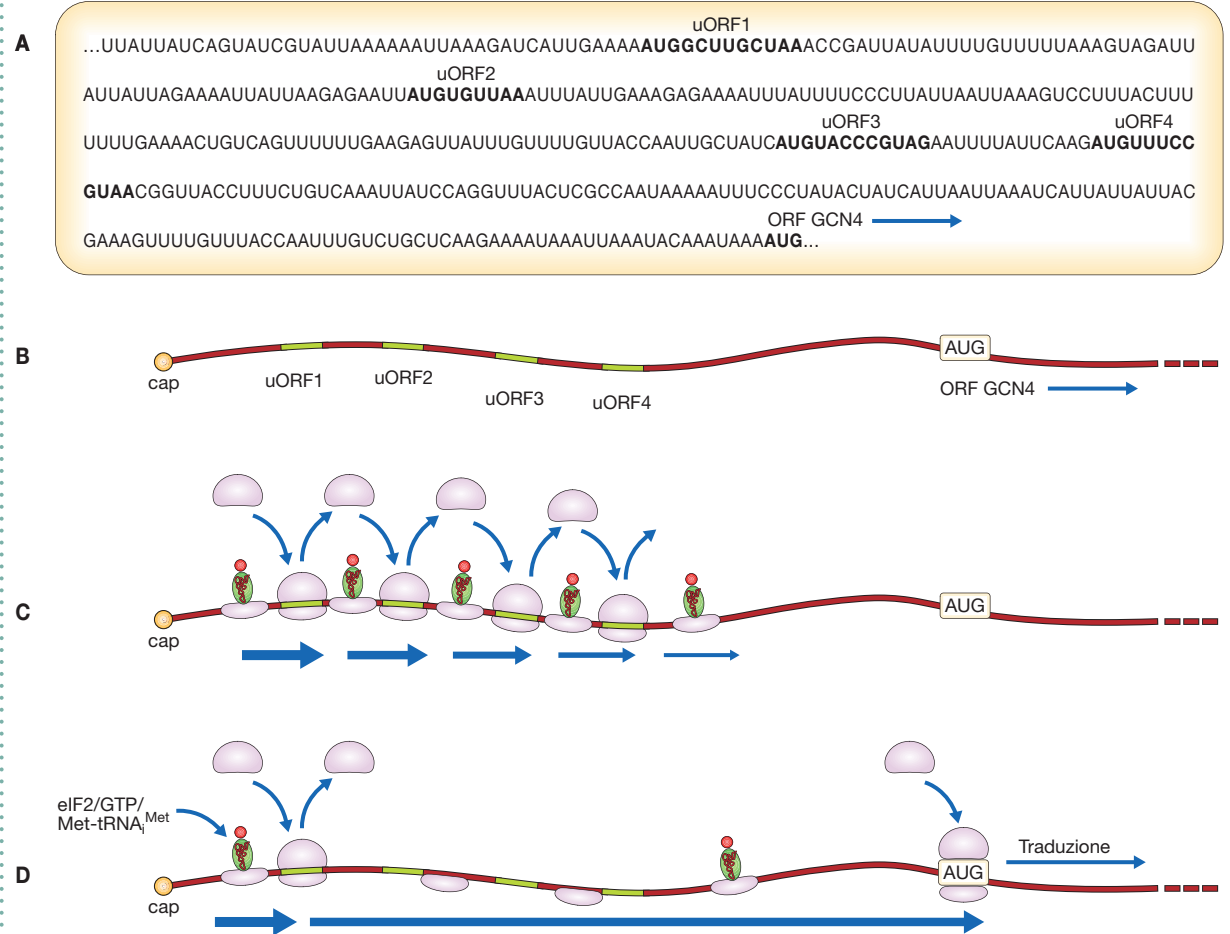


Figura 16.7 Controllo traduzionale del gene *GCN4* di lievito. **(A)** Sequenza nucleotidica della 5'UTR dell'mRNA *GCN4* di lievito in cui si vede la presenza di quattro brevissimi uORF. **(B)** Rappresentazione schematica della 5'UTR. **(C)** In presenza di alti livelli di aminoacidi e di complesso ternario, la subunità 40S riesce a reiniziare la traduzione su ciascuna delle 4 uORF, ma poiché ogni reinizio comporta una perdita del 50% dei ribosomi, pochissimi ar-

A ogni passaggio da una uORF alla successiva si ha una diminuzione della percentuale di subunità 40S che proseguono nella scansione, e quindi una bassa probabilità che raggiungano il vero AUG di inizio della regione codificante per GCN4. Invece, come mostrato in **figura 16.7D**, in condizioni di carenza di amminoacidi questa possibilità aumenta per due motivi: da una parte, come abbiamo visto precedentemente e in **figura 16.2**, la carenza di amminoacidi causa, tramite una via di trasduzione del segnale, la fosforilazione del fattore eIF2 e di conseguenza una riduzione della sua efficienza; dall'altra, la stessa carenza di amminoacidi risulta in un limitato livello di Met-tRNA_i di inizio. Queste due condizioni fanno sì che la subunità piccola che esce dalla uORF1 non trovi facilmente un complesso ternario Met-tRNA_i^{Met}/eIF2/GTP per potere reiniziare sulle uORF seguenti, le salti e continui invece la scansione. Poiché il vero AUG di inizio della regione codificante per GCN4 si trova più a valle notevolmente distanziato, la subunità 40S ha il tempo per reclutare un complesso ternario, anche se presente in quantità limitante, e può quindi iniziare la traduzione di GCN4.

16.3 Regolazione della stabilità e degradazione degli mRNA

Diverse classi di RNA cellulare sono caratterizzate in genere da diversa stabilità nel tempo. Per esempio, sia nei procarioti che negli eucarioti, i ribosomi, e quindi anche gli rRNA che ne fanno parte integrante, sono componenti molto stabili della cellula e una volta sintetizzati vengono riutilizzati per tempi lunghi relativamente al ciclo vitale della cellula stessa. Anche i tRNA sono molecole molto stabili. Al contrario, gli mRNA hanno in genere una vita media piuttosto breve, anche se non uguale per tutti. Nei procarioti alcuni mRNA hanno una vita media di pochi minuti, anche uno o due, altri di poco più lunga. Nelle cellule eucariotiche, dove il ciclo cellulare è 10-100 volte più lungo rispetto ai procarioti, anche gli mRNA hanno una vita media più lunga, ma sempre molto meno degli rRNA e tRNA.

Perché l'mRNA è così instabile? Un primo motivo può essere ricercato nella sua stessa funzione: l'mRNA non può essere permanentemente strutturato e assemblato in complessi ribonucleoproteici, come avviene per esempio per l'rRNA nei ribosomi, ma deve essere steso a singolo filamento per potere essere letto dai ribosomi durante la traduzione. L'assenza di una compatta struttura secondaria e di un'associazione a proteine può rendere le molecole di mRNA particolarmente soggette a rotture, che avvierebbero poi le molecole alla degradazione. Un altro motivo per cui gli mRNA hanno vita media relativamente breve è che a questa è correlata la velocità con cui una cellula può cambiare il suo programma sintetico. Infatti, solo con mRNA instabili si ottiene che le proteine sintetizzate corrispondano ai geni che sono in quel momento attivi. Se un mRNA fosse molto stabile, il suo prodotto proteico continuerebbe a essere sintetizzato anche dopo che il corrispondente gene è stato represso.

Degradazione dell'mRNA nei procarioti

Nei procarioti la degradazione degli mRNA comporta l'attività di varie endonucleasi ed esonucleasi e può iniziare quando ancora l'mRNA è in corso di traduzione. Come illustrato in **figura 16.8**, il processo inizia con tagli endonucleolitici e procede complessivamente in direzione 5'→3' dietro l'ultimo ribosoma, producendo numerosi frammenti. Questi vengono poi degradati completamente a singoli nucleotidi da esonucleasi del tipo 3'→5'. In *E. coli* ci sono circa 12 ribonucleasi implicate nei processi di maturazione e di degradazione degli RNA, in alcuni casi con funzioni in parte sovrapposte. Alcune ribonucleasi formano un complesso multienzimatico chiamato **degradosoma** che comprende l'endonucleasi RNasi E che produce il taglio iniziale, un'elicasi che svolge le strutture secondarie del frammento e la esonucleasi 3'→5' PNPasi che lo degrada a singoli nucleotidi.

Degradazione dell'mRNA nelle cellule eucariotiche

La regolazione della stabilità e degradazione degli mRNA negli eucarioti è alquanto più complessa ed è implicata in vari processi. Innanzitutto, la presenza nella cellula di esonucleasi 5'→3' e 3'→5' porta rapidamente a degradazione qualsiasi RNA le cui estremità non siano in qualche maniera protette. Nel caso degli mRNA questi sono protetti, come già abbiamo discusso precedentemente (*vedi fig. 14.16*), ad ambedue le estremità: il cappuccio (cap) protegge l'estremità 5' dall'attacco di esonucleasi 5'→3', mentre la coda di poli(A), con legate varie unità di PABP, protegge l'estremità 3' dall'attacco di esonucleasi 3'→5'. In questa maniera un mRNA integro è relativamente stabile. Se però subisce accidentalmente una rottura, i due frammenti che ne risultano vengono rapida-

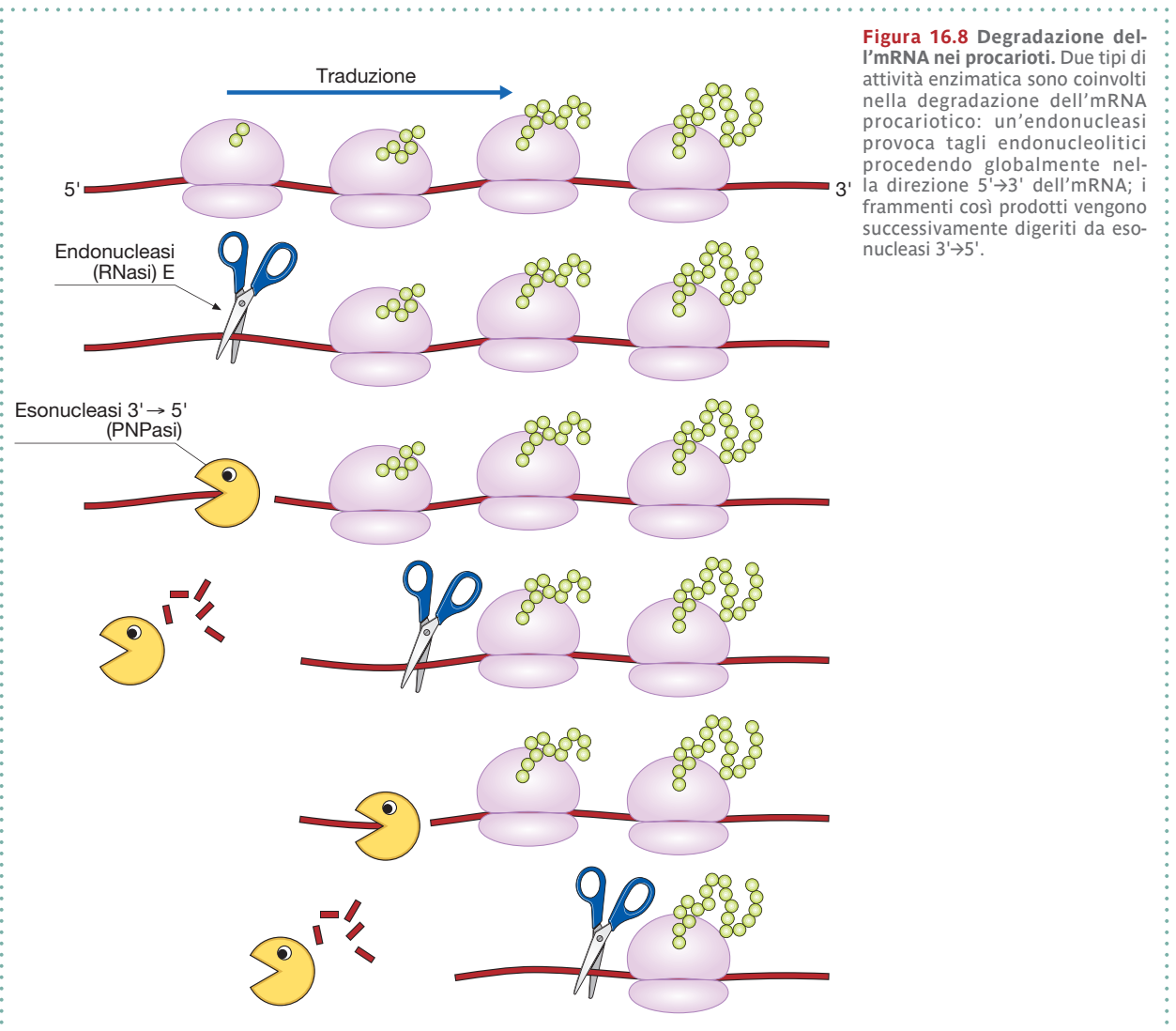


Figura 16.8 Degradazione dell'mRNA nei procarioti. Due tipi di attività enzimatica sono coinvolti nella degradazione dell'mRNA procariotico: un'endonucleasi provoca tagli endonucleolitici procedendo globalmente nella direzione 5'→3' dell'mRNA; i frammenti così prodotti vengono successivamente digeriti da esonucleasi 3'→5'.

mente degradati dalle esonucleasi (**fig. 16.9**). Questo meccanismo non specifico permette una continua rimozione dalla cellula degli mRNA danneggiati, il cui accumulo potrebbe essere dannoso.

D'altra parte, in accordo con l'idea che la degradazione specifica di un mRNA o la modulazione della sua degradazione possa rappresentare un importante meccanismo di regolazione dell'attività di alcuni geni negli eucarioti, è stata osservata anche la degradazione specifica di molecole di mRNA integre; diversi mRNA hanno differenti stabilità o tempi di degradazione a seconda del ruolo fisiologico delle proteine che codificano. Inoltre, la vita media di un particolare mRNA può essere in alcuni casi modulata in risposta a diverse condizioni o segnali cellulari. È stato dimostrato che le differenze nella velocità di degradazione degli mRNA sono determinate da elementi di sequenza presenti negli mRNA stessi. Per esempio, un segnale di destabilizzazione ben caratterizzato è l'elemento **ARE** (**AU-Rich Element**), la cui sequenza consenso è AUUUA ripetuta più volte nella 3'UTR di vari mRNA instabili (**fig. 16.10**). La presenza di questa sequenza innesca inizialmente la deadenilazione dell'mRNA, cioè la perdita della coda di poli(A) e quindi anche delle proteine PABP a essa legate, e porta alla degradazione dell'mRNA stesso (**fig. 16.10A**). In particolari condizioni le sequenze ARE possono essere legate da specifiche proteine, che così influenzano la stabilità del-

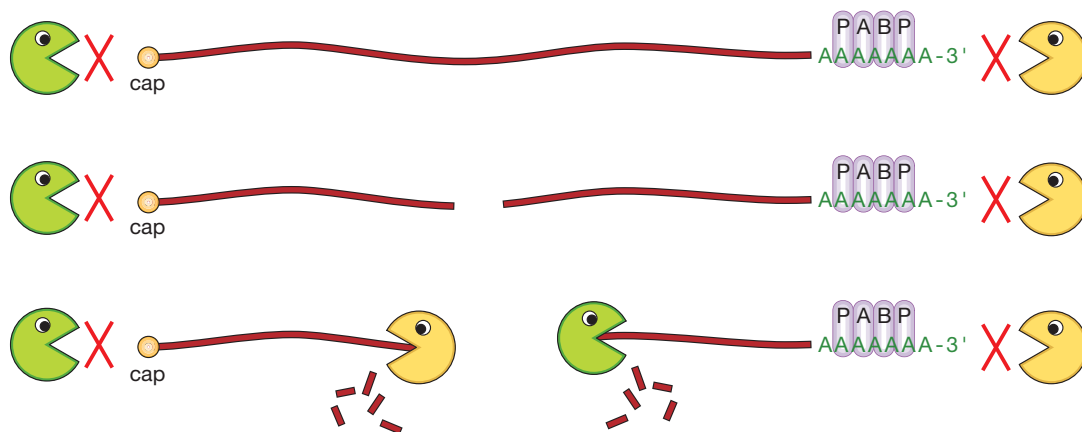


Figura 16.9 Degradazione degli mRNA danneggiati negli eucarioti. Il cappuccio al 5' e la PABP legata alla coda di poli(A) al 3' proteggono l'mRNA eucariotico dall'attacco delle esonucleasi

5'→3' e 3'→5'. Se l'mRNA subisce una rottura, esso viene rapidamente degradato dalle esonucleasi alle estremità 5' e 3' non protette che si sono create.

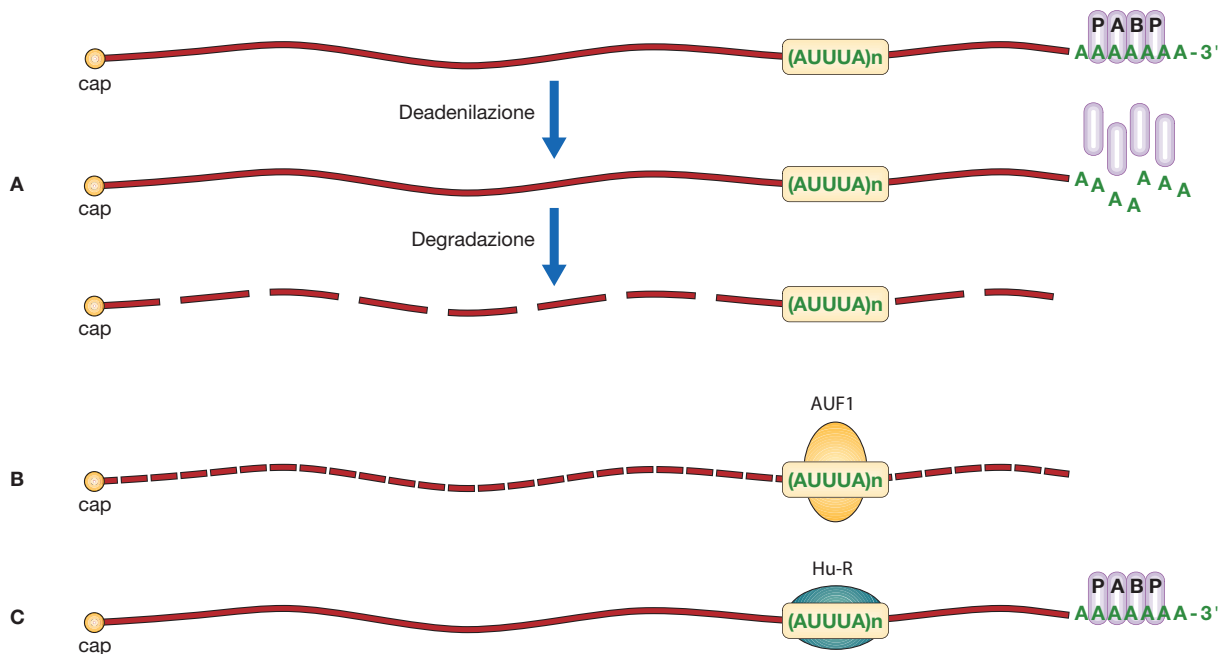
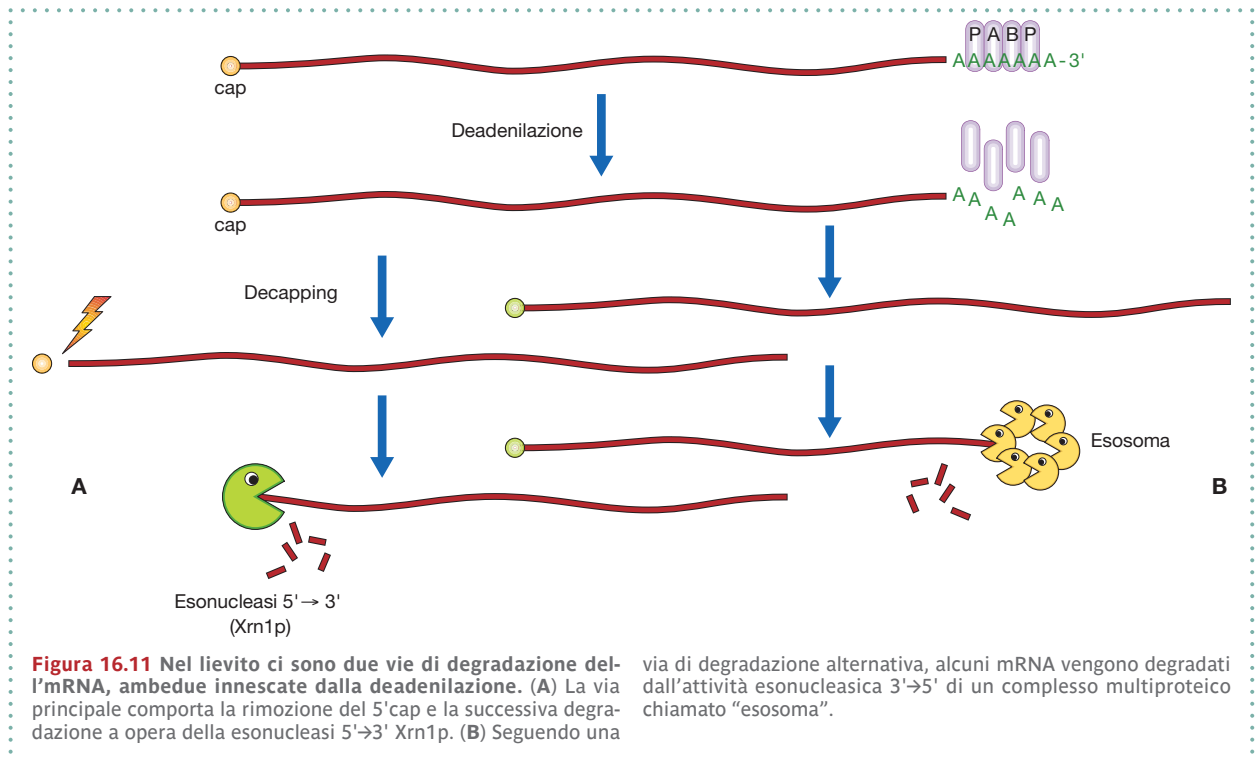


Figura 16.10 Sequenze ARE e controllo della stabilità degli mRNA. (A) Una sequenza ARE nella 3'UTR di un mRNA lo rende instabile accelerandone la deadenilazione e quindi la degradazione. (B) La proteina AUF1 lega la sequenza ARE che si trova nella

3'UTR dell'mRNA c-myc, accelerandone la degradazione. (C) La proteina Hu-R lega l'elemento ARE che si trova nella 3'UTR di vari mRNA bersaglio, stabilizzandoli.

l'mRNA. Per esempio, la proteina AUF1 umana lega l'elemento ARE presente nella 3'UTR dell'mRNA c-myc accelerandone la degradazione (**fig. 16.10B**). Al contrario, la proteina Hu-R lega l'elemento ARE di vari mRNA, tra cui quelli delle citochine, stabilizzandoli (**fig. 16.10C**). L'effetto stabilizzante di HuR è probabilmente dovuto alla competizione, con proteine destabilizzanti, per il legame con l'elemento ARE. Queste proteine possono quindi essere considerate fattori che agiscono in *trans* sui *cis*-elementi ARE per la regolazione della stabilità degli mRNA bersaglio.

Quali sono gli enzimi coinvolti in questo processo di degradazione degli mRNA eucariotici? Nel lievito, che è il sistema meglio studiato, ci sono due



via di degradazione dell'mRNA, illustrate in **figura 16.11**. Ambedue vengono inizialmente innescate dalla deadenilazione, cioè dalla rimozione della coda di poli(A), da parte di specifiche poli(A) nucleasi. Successivamente, seguendo la via principale di degradazione, si ha la rimozione del 5' cap da parte di un complesso che comprende gli enzimi di "Decapping" Dcp12/Dcp2 (**fig. 16.11A**). A questo punto l'mRNA, non più protetto dal cap, viene progressivamente degradato dalla esonucleasi 5'→3' Xrn1p. Oltre a questa via principale, è stata messa in evidenza una via secondaria di degradazione dell'mRNA (**fig. 16.11B**) per cui, dopo la deadenilazione, alcuni mRNA possono divenire substrato per l'esosoma, un complesso multiproteico con attività esonucleasica 3'→5'.

Meccanismi di "controllo qualità" dell'mRNA

Esistono sia nelle cellule procariotiche che in quelle eucariotiche numerosi sistemi di "sorveglianza" che assicurano, a vari livelli, un'elevata accuratezza dell'espressione genica. In particolare, vari meccanismi di "controllo qualità" degli mRNA portano a degradazione i messaggeri "difettosi" che vengono a formarsi nella cellula. Descriveremo in dettaglio il sistema NMD che, nelle cellule eucariotiche, porta a degradazione mRNA che presentino un codone di stop prematuro; accenneremo poi brevemente a due diversi meccanismi, NSMD e tmRNA, che operano, rispettivamente, negli eucarioti e nei procarioti per portare a degradazione mRNA che, al contrario, mancano del normale codone di terminazione.

L'**NMD** (*Non-sense Mediated Decay*) ha la funzione di eliminare gli mRNA che, contenendo nella regione codificante un surrettizio codone nonsense originatosi per mutazione o per errore di trascrizione o di splicing, codificherebbero proteine incomplete all'estremità C-terminale. Infatti, spesso è meno dannoso non produrre affatto una proteina piuttosto che produrne una difettosa. Poiché il sistema funziona per la maggior parte degli mRNA che codificano proteine diverse, non può essere la particolare struttura dell'mRNA stesso o della proteina da esso codificata a segnalare che quel particolare mRNA deve essere degradato.

Ci deve essere piuttosto un sistema che permette di identificare se un codone di stop è quello naturale che si trova alla fine della regione codificante (ORF) o se è un codone di stop prematuro (chiamato alle volte PTC, *Premature Termination Codon*), cioè un codone di stop dovuto a una mutazione o a un errore di trascrizione o di splicing.

Il sistema è stato studiato in lievito, in *C. elegans* e anche negli animali superiori. Sono state identificate alcune proteine (per esempio Upf1 e Upf2 negli animali superiori) che in certe condizioni innescano la degradazione dell'mRNA a cui

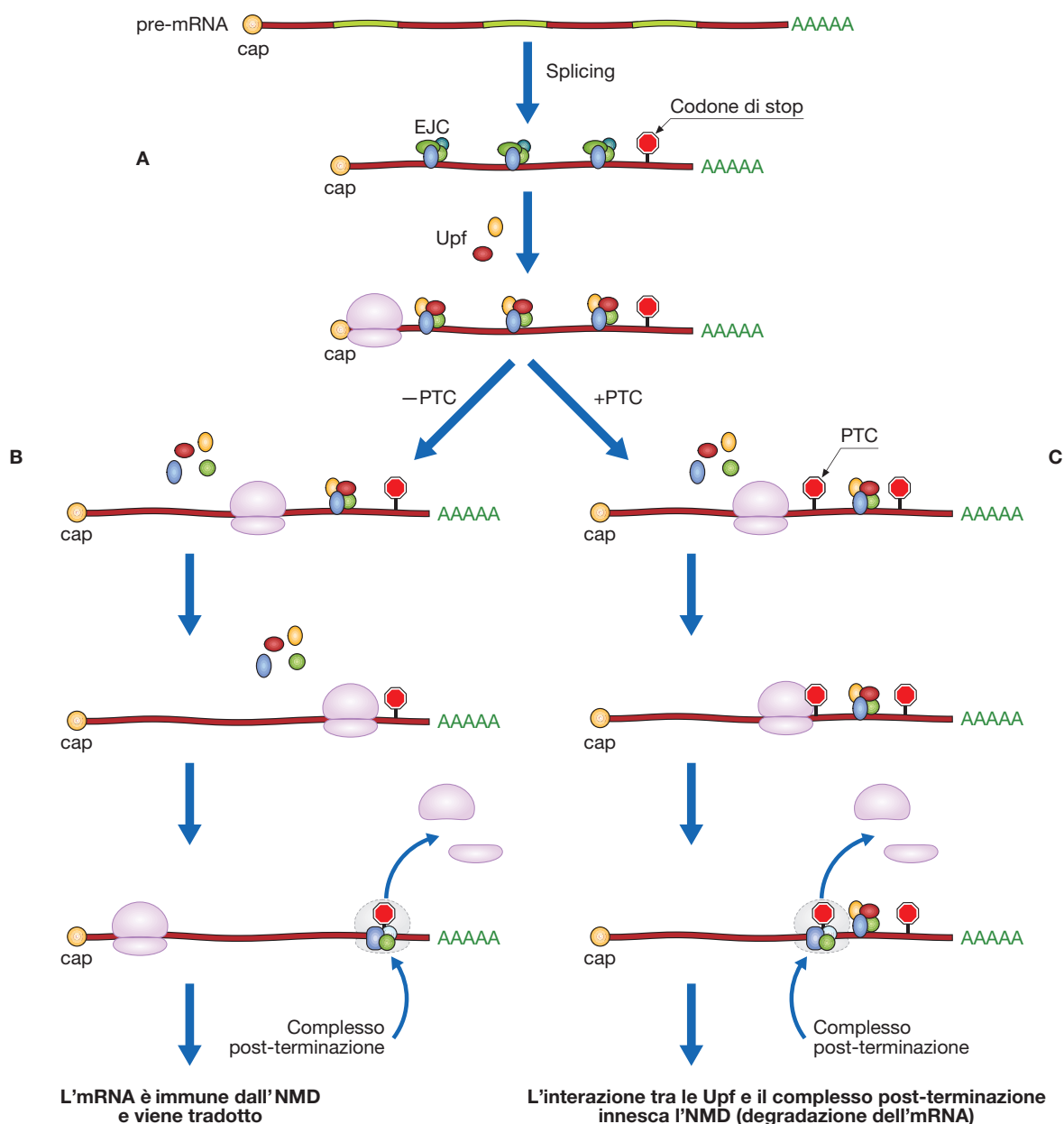


Figura 16.12 Il sistema di sorveglianza NMD (*Non-sense Mediated Decay*) degli eucarioti. (A) Dopo lo splicing i siti dell'mRNA da cui sono stati rimossi gli introni restano marcati dalla presenza di complessi proteici EJC. Questi subiscono poi dei cambiamenti strutturali, tra cui il legame delle proteine Upf, che, se non rimosse, porteranno l'mRNA a degradazione. (B) Al

primo ciclo di traduzione il ribosoma che scorre su un mRNA normale rimuove le proteine Upf rendendo così l'mRNA stabile. (C) Se l'mRNA contiene un codone di stop prematuro (PTC), questo causa il rilascio del ribosoma prima che esso abbia raggiunto e rimosso la seguente proteina Upf, che quindi causerà la degradazione dell'mRNA.

sono legate e, in alcuni casi, è stato trovato che in qualche modo l'NMD è collegato al processo di splicing. Il meccanismo molecolare alla base dell'NMD non è stato ancora completamente chiarito, ma sono stati proposti vari modelli di cui descriviamo qui quello indicato per cellule di mammifero (**fig. 16.12**). Come è stato descritto nel paragrafo 13.1 e come è mostrato in **figura 16.12A**, ciascun sito dell'mRNA da cui è stato rimosso un introne resta marcato, dopo il disassemblaggio dello spliceosoma, dalla presenza di un complesso proteico, chiamato **EJC**, che resta legato in prossimità della giunzione esone-esone. La presenza di questi complessi EJC legati ai vari siti di splicing lungo un mRNA, e l'assenza di componenti dello spliceosoma, segnalano il fatto che l'mRNA è ormai maturo. I complessi EJC subiscono poi dei cambiamenti strutturali con perdita di alcune proteine e legame di altre, che restano a marcare le giunzioni esone-esone. Alcune di queste sono implicate nell'esportazione nucleo-citoplasma dell'mRNA. Altre, Upf1 e Upf2, sono implicate nel sistema di sorveglianza NMD. Infatti, queste proteine sono responsabili nel citoplasma della degradazione dell'mRNA, a meno che non vengano rapidamente rimosse dal passaggio del primo ribosoma che legge l'mRNA stesso. La **figura 16.12B** mostra come in condizioni normali, cioè in assenza di codoni di stop prematuri, il ribosoma rimuove i complessi Upf rendendo così l'mRNA immune al NMD. La **figura 16.12C** mostra invece come la presenza di un codone di stop prematuro, a monte dell'ultimo introne, fa sì che l'ultimo complesso Upf non venga rimosso dal ribosoma e la sua interazione con il complesso di post-terminazione porta alla degradazione dell'mRNA. È per questo forse che in genere non ci sono introni nella regione che codifica per la 3'UTR; infatti, in questo caso il codone di stop naturale sarebbe interpretato dal sistema come un codone di stop prematuro. Ribadiamo che questo modello, pur essendo in accordo con varie osservazioni sperimentali, non è ancora dimostrato e rappresenta solo uno dei possibili modelli proposti per l'NMD.

Come abbiamo già accennato, sia procarioti che eucarioti hanno sistemi di controllo qualità per eliminare mRNA che mancano del normale codone di terminazione alla fine della regione codificante. Poiché l'assenza di codoni di stop su mRNA mutati o tronchi può portare allo stallo dei ribosomi all'estremità 3' del

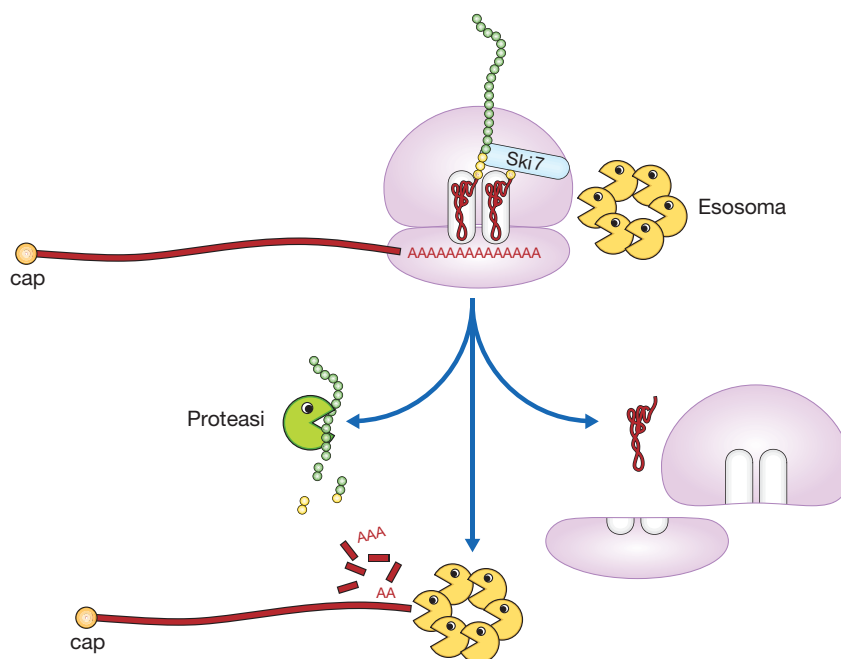
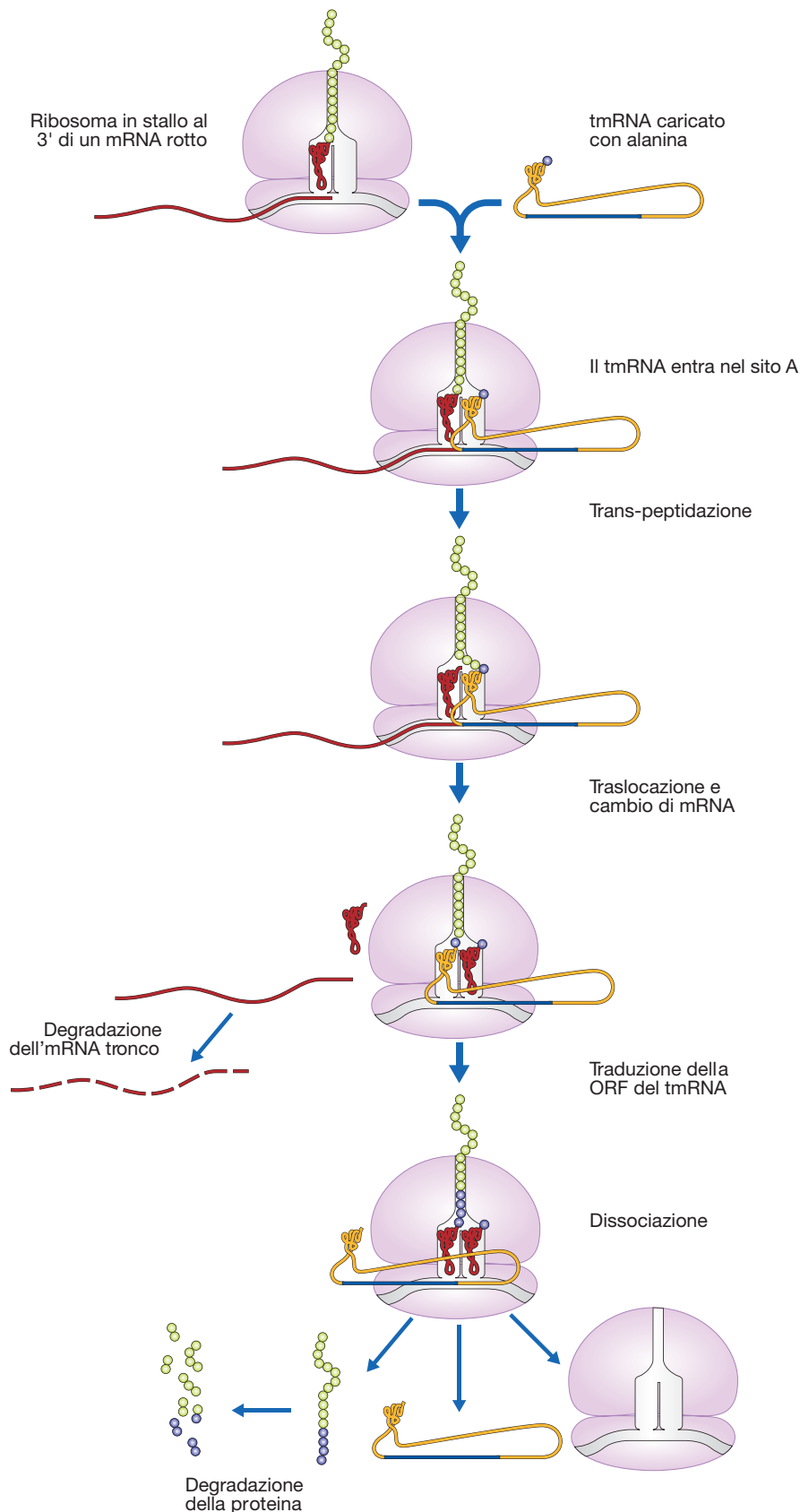


Figura 16.13 Il sistema di sorveglianza NSMD (*Non-Stop Mediated Decay*) degli eucarioti. In assenza di un codone di terminazione il ribosoma prosegue la traduzione sulla coda di poli(A) aggiungendo al peptide nascente (in verde) una serie di lisine (in giallo). Questo tratto di poli-lisine recluta la proteina Ski7, affine al fattore di rilascio RF3, che stimola il rilascio del ribosoma e l'attivazione del complesso dell'esosoma che degrada l'mRNA. Anche la proteina sintetizzata, che termina con un tratto di poli-lisine, viene indirizzata alla degradazione.

Figura 16.14 Ruolo del tmRNA dei procarioti. Nei procarioti l'assenza del codone di terminazione provoca lo stallo del ribosoma all'estremità 3' dell'mRNA. Per sbloccare questa situazione e avviare l'mRNA difettoso alla degradazione, interviene il tmRNA (arancione), le cui estremità formano una struttura simile a un tRNA e che è caricato con alanina, mentre nella parte centrale presenta una breve ORF (blu) di 10 codoni. Il tmRNA entra nel sito A del ribosoma in stallo e permette la prosecuzione della traduzione: infatti, esso offre l'alanina su cui viene trasferito il peptide nascente e, al tempo stesso, inserisce la sua ORF nel canale dell'mRNA estromettendo l'mRNA tronco, che viene rapidamente degradato. La traduzione della ORF del tmRNA risulta nell'aggiunta al peptide nascente di una sequenza di 10 amminoacidi ricca in alanine, che sono un segnale per la sua degradazione.



messaggero, sono necessari meccanismi che permettano di dissociare il ribosoma dall'mRNA ed eventualmente avviare questo alla degradazione. Negli eucarioti è attivo un meccanismo, chiamato **NSMD** (*Non-Stop Mediated Decay*) e illustrato in **figura 16.13**, che porta a degradazione l'mRNA quando, a causa di una terminazione prematura della trascrizione o di una mutazione, non ha il codone di stop. Dato che questo mRNA è comunque poliadenilato, la traduzione continua sulla coda di poli(A) producendo un tratto di poli-lisine. Come illustrato in figura, questo tratto poli-lisिनico è un segnale per il reclutamento della proteina Ski7, simile al fattore RF3, che causa il rilascio del ribosoma e avvia alla degradazione l'mRNA difettoso e il peptide sintetizzato che ha l'estremità C-terminale anormale.

Nei procarioti, dove l'mRNA non è poliadenilato, la dissociazione dall'mRNA del ribosoma in stallo all'estremità 3' è resa possibile dall'intervento di un RNA molto particolare chiamato **tmRNA** (o anche **SsrA**) perché funziona sia da tRNA che da mRNA. Diversi organismi presentano diverse varianti dello stesso meccanismo; nel caso di *E. coli* il tmRNA, lungo circa 450 nt, ha le estremità appaiate e strutturate in maniera simile a un tRNA ed è capace di caricare alanina, mentre nella parte centrale della molecola è presente una breve ORF che codifica per una decina di amminoacidi, tra cui varie alanine, e termina con un codone di stop. Come mostrato in **figura 16.14**, il tmRNA caricato con alanina (legato a EF-Tu-GTP) entra nel sito A del ribosoma che si trova in stallo all'estremità 3' dell'mRNA e posiziona la regione della ORF nell'apposito canale nel ribosoma espellendo l'mRNA tronco, che viene rapidamente degradato. La traduzione riprende allungando il peptide già sintetizzato con gli amminoacidi codificati dalla ORF del tmRNA, fino a che non viene raggiunto il codone di stop sul tmRNA stesso. La presenza di alanine all'estremità della proteina sintetizzata rappresenta un segnale per la degradazione da parte di specifiche proteasi.

FINESTRA 16.1

PATOLOGIE DELL'APPARATO DI TRADUZIONE

Poiché la sintesi proteica rappresenta uno dei processi di base essenziali per la vita delle cellule e degli organismi, ci si aspetta che difetti genetici dell'apparato di traduzione abbiano conseguenze molto gravi, in genere letali. Tuttavia ci possono essere situazioni intermedie di difetti abbastanza limitati che, pur creando problemi, restano compatibili con la sopravvivenza. Infatti sono noti mutanti ribosomali (di geni per rRNA o per r-proteine) nel lievito *Saccharomyces cerevisiae*, in *Drosophila* e in *Xenopus laevis*, i cui fenotipi comprendono ridotta crescita, ritardo dello sviluppo e una serie di anomalie strutturali e funzionali dell'organismo. Per quanto riguarda l'uomo sono state recentemente osservate alcune correlazioni tra specifiche situazioni patologiche e alcuni geni coinvolti nella produzione e funzione dell'apparato traduzionale. Per esempio:

1. il 25% dei malati di anemia di Diamond-Blackfan (DBA), che comporta aplasia dei globuli rossi causata da difetti intrinseci dei precursori eritropoietici, hanno mutazioni nel gene per la proteina ribosomale S19;
2. malati di discheratosi congenita (DC), rara sindrome ereditaria che comporta insufficienza del midollo osseo e varie anomalie associate, presentano mutazioni del gene per la discherina, proteina nucleolare coinvolta nella biogenesi del ribosoma oltre che nel mantenimento dei telomeri cromosomici;

3. mutazioni delle subunità di eIF2B sono alla base della CACH (o VWM), una forma di atassia infantile con ipomielinizzazione del sistema nervoso;
4. mutazioni del gene per la proteina nucleolare nucleophosmina (chiamata anche NPM, B23 e NO38) risultano in una severa anemia conseguente a difetti dell'organogenesi precoce.

Si può notare come disordini ematologici e anemia ricorrano spesso nei casi di mutazioni di componenti dell'apparato traduzionale. Ciò suggerisce che l'ematopoiesi comporti una massima richiesta di sintesi proteica e che quindi sia la prima a essere colpita quando l'apparato di traduzione è deficitario.

Inoltre, poiché una corretta regolazione della crescita e proliferazione cellulare non può prescindere da una corretta regolazione della sintesi proteica globale, non è da stupirsi che siano state fatte numerose osservazioni che indicano che difetti della traduzione e della sua regolazione siano implicati in processi di cancerogenesi.

Anche se le informazioni sono ancora molto scarse, si possono prevedere interessanti sviluppi in cui la conoscenza di base dei meccanismi di traduzione permetterà di capire alcune situazioni patologiche, mentre queste ultime daranno materiale utile per allargare le nostre conoscenze di base.

16.4 Trasporto e localizzazione degli mRNA

Nelle cellule procariotiche gli RNA svolgono la loro funzione nell'unico compartimento cellulare, lo stesso dove vengono sintetizzati, tanto che in molti casi gli mRNA possono iniziare a essere tradotti quando ancora non è terminata la loro trascrizione sullo stampo di DNA. Al contrario, nelle cellule eucariotiche gli RNA, dopo essere stati sintetizzati, devono essere trasportati da un compartimento all'altro della cellula e, alle volte, devono localizzarsi in particolari regioni del compartimento in cui si vanno a trovare.

Tra i processi di trasporto attraverso le membrane che delimitano i vari compartimenti delle cellule eucariotiche troviamo innanzitutto il passaggio dal nucleo al citoplasma degli RNA che costituiscono l'apparato di traduzione, rRNA, tRNA e mRNA. Inoltre, alcuni RNA prodotti nel nucleo ed esportati nel citoplasma vengono ulteriormente spostati all'interno dei mitocondri o dei cloroplasti, come per esempio avviene per alcuni tRNA implicati nella traduzione mitocondriale. Questi trasporti richiedono spesso meccanismi specifici per superare intrinseche difficoltà, come l'attraversamento di membrane idrofobiche da parte di molecole con forte carica negativa, quali sono gli RNA. In genere, il trasporto non riguarda la molecola di RNA isolata, ma assemblata con proteine specifiche. Per esempio, gli rRNA escono dal nucleo al citoplasma in forma di subunità ribosomali in cui sono complessati alle r-proteine, i tRNA vengono trasportati da un sistema di proteine specifiche e anche gli mRNA escono dal nucleo al citoplasma, in forma di mRNP, cioè complessi ribonucleoproteici formati da mRNA e proteine associate. Una volta raggiunto il citoplasma, alle volte gli RNA devono andare a localizzarsi in regioni specifiche di questo. Si ottiene così una distribuzione asimmetrica di mRNA nella cellula, che spesso è essenziale per il corretto sviluppo embrionale o per stabilire asimmetrie strutturali e funzionali della cellula, come nel tipico esempio dei neuroni. La specificità della localizzazione citoplasmatica asimmetrica è data dalla presenza nell'mRNA di specifiche sequenze *cis*-agenti che vengono riconosciute da proteine che legano l'RNA (*RNA-binding*). Il risultante complesso mRNA-proteine (mRNP) si lega direttamente o indirettamente a un motore molecolare che lo trasporta al sito di destinazione finale, dove l'mRNA viene in qualche modo ancorato e poi tradotto. Spesso esistono meccanismi che impediscono la traduzione prematura di un mRNA durante il trasporto. Le sequenze *cis*-agenti che individuano l'mRNA che deve essere trasportato vengono da alcuni autori chiamate **zip-code** perché indicano la destinazione finale di un mRNA (*zip-code* in inglese corrisponde al nostro CAP = codice di avviamento postale). Le sequenze zip-code possono essere più o meno lunghe, da poche decine a qualche centinaia di nucleotidi. Esse si trovano perlopiù nella regione 3'UTR dell'mRNA, soprattutto nei metazoi, ma alle volte si possono trovare anche nella regione codificante (ORF), come avviene per alcuni mRNA di lievito. Numerosi esempi di questo tipo di processi si trovano in organismi eucariotici, dal lievito ai mammiferi. Qui ne illustreremo tre: il caso relativamente semplice e ben studiato dell'mRNA di ASH1 di lievito, gli mRNA materni coinvolti nella polarizzazione dell'uovo di *Drosophila* e la localizzazione di alcuni mRNA nei dendriti di cellule neuronali.

Localizzazione dell'mRNA di ASH1 in lievito

Saccharomyces cerevisiae è un lievito gemmante che, al momento della duplicazione cellulare, forma una gemma che poi si accresce dando origine alla cellula figlia. Solo in un secondo tempo il nucleo della cellula madre si divide per mitosi, e uno dei due nuclei prodotti migra nella cellula figlia. Il prodotto proteico del gene *ASH1* svolge la sua funzione all'estremità della gemma (cellula figlia) quando questa non ha ancora un suo nucleo. Pertanto, l'mRNA di ASH1, trascritto dal

gene presente nel genoma della cellula madre, dopo essere esportato dal nucleo al citoplasma deve migrare nella gemma, dove infine viene tradotto. Trattandosi di lievito, è stato possibile dissezionare il sistema mediante un approccio genetico che ha permesso di individuare cinque geni, *SHE1-SHE5*, implicati nel processo di localizzazione dell'mRNA. Infatti, mutazioni in uno qualsiasi di questi geni impediscono la migrazione e localizzazione specifica dell'mRNA di *ASH1*. Le proteine She1, She2 e She3 formano un complesso ribonucleoproteico con l'mRNA di *ASH1* che viene trasportato nella cellula figlia. Come mostrato in **figura 16.15**, She2 riconosce e lega una sequenza zip-code presente nell'mRNA di *ASH1*, mentre She3 connette She2 a She1. Quest'ultima, chiamata anche Myo4, è una forma particolare di miosina. Come è noto, le miosine sono dei "motori

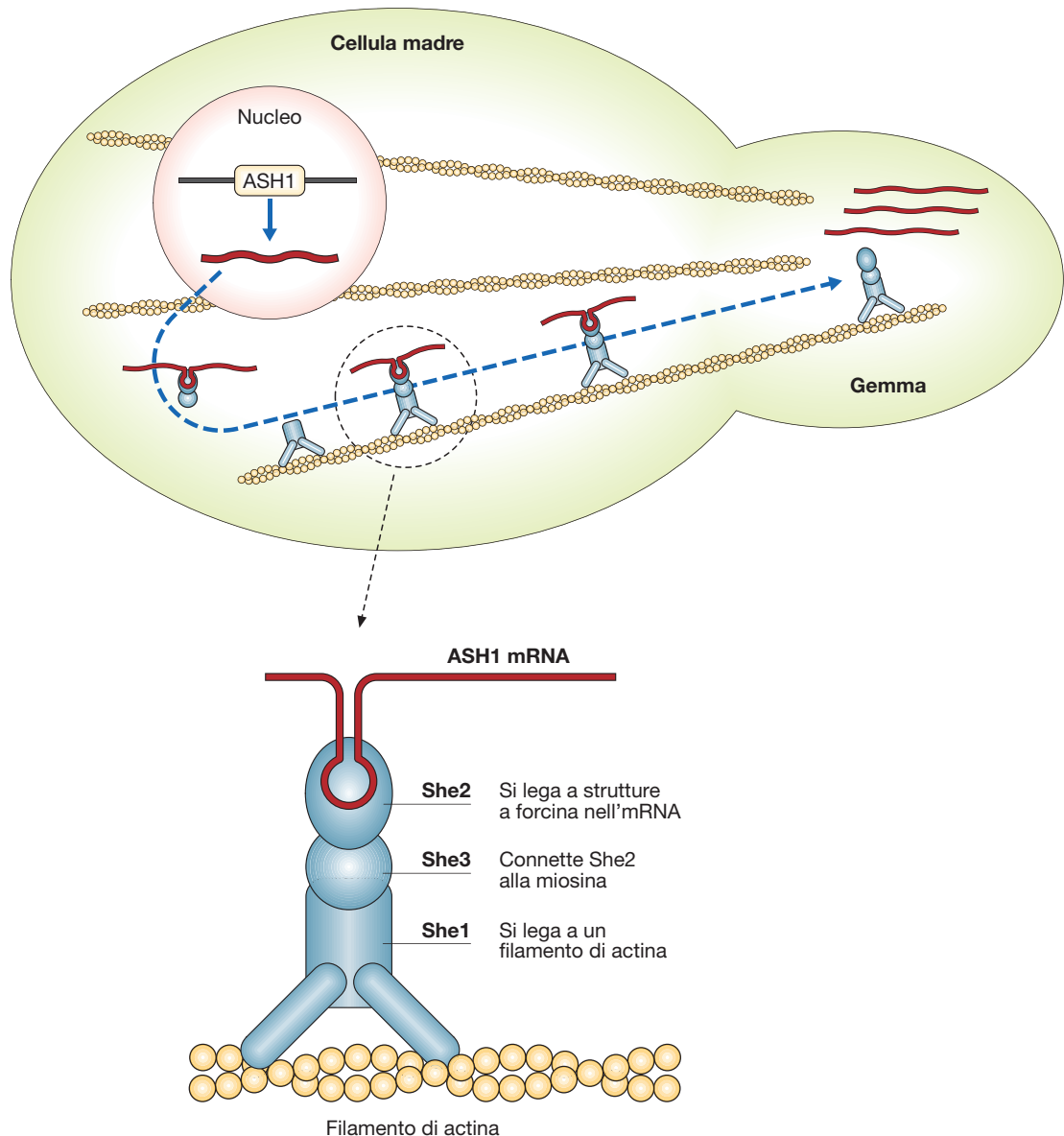


Figura 16.15 Trasporto e localizzazione dell'mRNA di *ASH1* dalla cellula madre alla cellula figlia nel lievito *Saccharomyces cerevisiae*. L'mRNA di *ASH1* trascritto nel nucleo della cellula madre viene trasportato nel citoplasma, dove forma un complesso

ribonucleoproteico con le proteine She. Questo complesso contiene un motore di miosina (She1) che lo sposta lungo i filamenti di actina trasportando così l'mRNA fino alla gemma. [Modificata da: Long, R.M. (2000), *The EMBO Journal*, 19, pp. 6592-6601.]

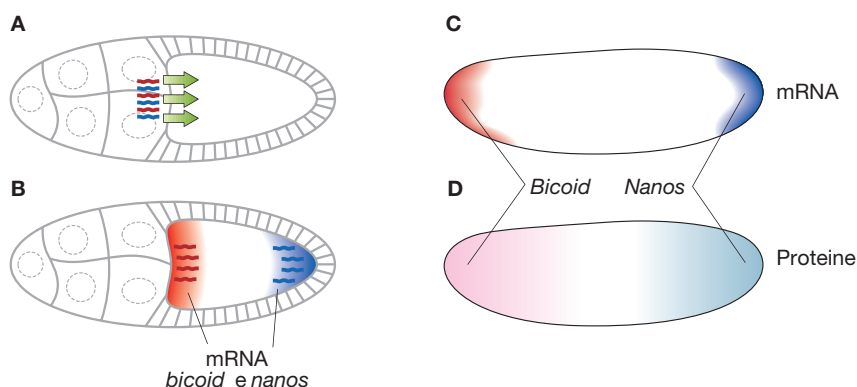
molecolari” capaci di spostarsi lungo i filamenti di actina, che costituiscono uno dei componenti del citoscheletro. Di conseguenza il complesso mRNP costituito dall'mRNA di ASH1 e dalle tre proteine She1, She2 e She3, si sposta lungo i filamenti polarizzati di actina che si estendono dal citoplasma della cellula madre al citoplasma della cellula figlia. Così l'mRNA di ASH1 raggiunge e si accumula nella cellula figlia, dove viene infine tradotto. In realtà questo mRNA contiene quattro sequenze zip-code, una nella 3'UTR e 3 nella regione codificante, caratterizzate da una particolare struttura a forcina e tutte necessarie per un completo ed efficiente trasporto dell'mRNA di ASH1.

Localizzazione di mRNA nell'uovo di *Drosophila*

La localizzazione di mRNA in regioni specifiche del citoplasma è un fenomeno diffuso e importante nello sviluppo embrionale di molti organismi in cui processi di questo tipo contribuiscono, per esempio, alla formazione dell'asimmetria lungo l'asse antero-posteriore, necessaria per un corretto sviluppo e differenziamento. Il caso più studiato è quello di *Drosophila*, dove troviamo sia il trasporto di mRNA dalle cellule nutrici all'oocita attraverso ponti citoplasmatici sia la localizzazione di vari mRNA in diverse specifiche regioni dell'oocita. A rendere il sistema più complesso, ma anche più interessante, è la commistione di processi di regolazione della localizzazione e della traduzione dei vari mRNA. Consideriamo in particolare il caso degli mRNA *nanos* e *bicoid* illustrato in **figura 16.16**. Questi fanno parte di un gruppo di vari mRNA prodotti per trascrizione di geni “materni” nei nuclei delle cellule nutrici e poi trasportati nell'oocita, attraverso dei ponti citoplasmatici (**fig. 16.17A**). Una volta entrati nell'oocita i vari mRNA materni hanno destini diversi e, in particolare, gli mRNA *nanos* e *bicoid* vanno a localizzarsi ai poli opposti dell'oocita. Infatti, l'mRNA *nanos* viene trasportato attraverso tutto il citoplasma dell'oocita fino al polo posteriore, mentre l'mRNA *bicoid* resta concentrato nel polo anteriore (**fig. 16.17B**). I due mRNA restano localizzati anche durante i successivi stadi di crescita dell'oocita e i primi stadi di sviluppo dell'embrione (**fig. 16.17C**), dove la traduzione localizzata di questi due mRNA, seguita da una certa diffusione dei prodotti della traduzione, risulta in due gradienti esponenziali, uno antero-posteriore della proteina bicoid e uno postero-anteriore della proteina nanos (**fig. 16.17D**). Queste proteine sono ambedue dei repressori traduzionali specifici che inibiscono, rispettivamente, la traduzione degli mRNA *hunchback* e *caudal* che, pur essendo distribuiti uniformemente nell'oocita, produrranno due gradienti di concentrazione contrapposti e con andamento sigmoide delle rispettive proteine.

Come avvengono il trasporto e la localizzazione degli mRNA *nanos* e *bicoid* ai due poli dell'oocita? È stato dimostrato che in questo sistema è implicato

Figura 16.16 Localizzazione citoplasmatica di mRNA durante lo sviluppo di *Drosophila*. (A) Alcuni mRNA sintetizzati nelle cellule nutrici, tra cui *bicoid* e *nanos*, vengono trasportati all'oocita attraverso canali citoplasmatici. (B) Successivamente gli mRNA *bicoid* e *nanos* vanno a localizzarsi ai due poli dell'oocita. (C) Gli mRNA *bicoid* e *nanos* restano localizzati durante i primi stadi di sviluppo dell'embrione. (D) La traduzione degli mRNA *bicoid* e *nanos* risulta in due gradienti contrapposti di concentrazione delle rispettive proteine.



l'altro componente del citoscheletro, cioè il motore molecolare dineina con la rete di microtubuli che attraversano anche i ponti citoplasmatici che uniscono i citoplasmi di cellule nutrici e dell'oocita. In particolare una proteina chiamata **Staufen** si lega a una sequenza (zip-code) che ha una particolare struttura e che si trova nella 3'UTR dell'mRNA *bicoid*. Il complesso mRNA *bicoid*/Staufen si associa con le estremità "meno" dei microtubuli, risultando così nella localizzazione dell'mRNA *bicoid* al polo anteriore dell'oocita. L'mRNA *nanos*, invece, migra verso le estremità "più" dei microtubuli, andandosi così a localizzare all'estremità posteriore. Questa è una descrizione molto semplificata del sistema: altre proteine sono coinvolte nella localizzazione e traduzione di questi due mRNA, e altri mRNA importanti per lo sviluppo vengono regolati da altri fattori proteici.

Localizzazione di mRNA nei dendriti neuronali

In genere la localizzazione di mRNA è molto evidente nelle cellule polarizzate; spesso la polarizzazione di una cellula dipende proprio dalla localizzazione di alcuni mRNA. I neuroni sono cellule estremamente polarizzate con un corpo cellulare da cui si dipartono due tipi di neuriti, l'assone e i dendriti. È stato osservato che alcuni mRNA vanno a localizzarsi in questi prolungamenti per portare avanti una sintesi di proteine in loco.

Per esempio, la **figura 16.17A** mostra la presenza dell'mRNA per α -CamKII nei prolungamenti dendritici di un neurone in coltura, mentre l'mRNA per la α -tubulina, usato come controllo negativo, è localizzato esclusivamente nel corpo cellulare.

Similmente, la **figura 16.17B** mostra che, mentre l'mRNA di controllo per l' α -tubulina è localizzato nello strato dell'ippocampo dove sono confinati i corpi cellulari dei neuroni, l'mRNA per la MAP2 è diffuso anche nella regione contenente solo prolungamenti neuronali. Per quanto riguarda il meccanismo di trasporto di questi mRNA è stato dimostrato, nel caso dell'mRNA per la α -CamKII, che la localizzazione dell'mRNA nei dendriti e alle sinapsi dipende da un tratto di oltre 1000 nt che si trova nell'ultima parte della lunga 3'UTR e che contiene tre sequenze zip-code; anche in questo caso, sarebbe coinvolta la proteina Staufen.

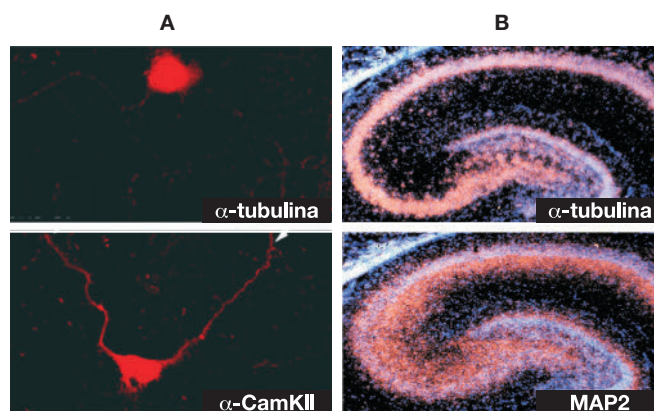


Figura 16.17 Localizzazione di mRNA nei dendriti neuronali messa in evidenza mediante ibridazione *in situ*. (A) L'ibridazione *in situ* con sonde specifiche mette in evidenza la presenza di mRNA per α -CamKII nei prolungamenti dendritici di un neurone in coltura, mentre l'mRNA per la α -tubulina usato come controllo negativo è localizzato esclusivamente nel corpo cellulare.

(B) L'ibridazione *in situ* su sezioni di ippocampo di ratto mostra che, mentre l'mRNA di controllo per l' α -tubulina è localizzato nello strato dell'ippocampo dove sono confinati i corpi cellulari dei neuroni, l'mRNA per la MAP2 sono diffusi anche nella regione contenente solo prolungamenti neuronali.

16.5 Granuli citoplasmatici di RNA

È stato recentemente osservato che nelle cellule eucariotiche i processi di regolazione della traduzione, stabilità e localizzazione degli mRNA di cui abbiamo parlato, sono in qualche maniera connessi alla localizzazione degli mRNA stessi in vari tipi di particelle cellulari, chiamate genericamente **granuli citoplasmatici di RNA** (*cytoplasmic RNA granules*). Vari tipi di granuli di RNA sono stati descritti in diversi tipi cellulari: PB (*Processing Bodies*) e SG (*Stress Granules*) nelle cellule somatiche, GCG (*Germinal Cell Granules*) nelle cellule germinali, granuli neuronali nei neuroni e altri. Tutti questi tipi di granuli contengono, oltre al loro carico di mRNA, anche subunità ribosomali, fattori di traduzione, enzimi nucleolitici, elicasi, proteine strutturali e proteine che legano l'RNA. Alcuni di questi componenti sono comuni ai vari tipi di granuli, mentre altri sono specifici per alcuni di essi. Descriveremo qui brevemente solo alcuni più importanti tipi di granuli di RNA.

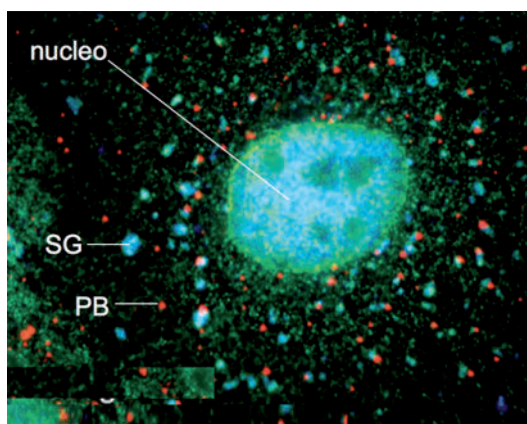
La **figura 16.18** mostra una fotografia di una cellula in cui si vedono PB e SG evidenziati al microscopio confocale mediante anticorpi fluorescenti che riconoscono proteine specifiche dell'una o dell'altra struttura.

I **PB** (*Processing Bodies*), studiati soprattutto in lievito e in cellule di mammifero, sono costitutivamente presenti e sarebbero coinvolti soprattutto nel controllo della stabilità/degradazione degli mRNA. Vi troviamo, infatti, localizzati alcuni componenti di vari sistemi di degradazione dell'mRNA: gli enzimi Dcp1p e Dcp2p per la rimozione del cappuccio, enzimi coinvolti nella degradazione dell'RNA come l'esonucleasi 5'→3', elementi coinvolti nel sistema di sorveglianza NMD, proteine componenti di RISC e altri ancora.

Gli **SG** (*Stress Granules*) compaiono nelle cellule di mammifero quando sottoposte a vari tipi di stress (stress da calore, stress ossidativo, radiazioni UV, ipossia ecc.). In genere la cellula sottoposta a stress riprogramma il proprio metabolismo degli mRNA per riparare i danni indotti dallo stress e per adattarsi alle nuove condizioni. Durante questo processo, la traduzione degli mRNA che codificano per le proteine housekeeping viene bloccata, mentre aumenta la traduzione di mRNA che codificano per proteine chaperone e per enzimi coinvolti nella riparazione dei danni provocati dallo stress. Gli mRNA che si trovano negli SG non vanno incontro a degradazione ma restano traduzionalmente inattivi o associati a subunità ribosomali 40S. Essi possono essere eventualmente rilasciati dai granuli e traduzionalmente riattivati.

I meccanismi di controllo della localizzazione, stabilità e traduzione degli mRNA di cui abbiamo trattato in questo capitolo, coinvolgono tipicamente in-

Figura 16.18 *Processing Bodies* e *Stress Granules*. SG e PB sono evidenziati mediante immunofluorescenza con anticorpi verso proteine specifiche per gli SG (blu) e i PB (rosso) in una cellula in cui è stato indotto stress ossidativo mediante trattamento con arsenio.



terazioni tra sequenze nucleotidiche presenti in *cis* negli mRNA bersaglio ed elementi in *trans* rappresentati da fattori proteici. Si è poi trovato che alcuni specifici tipi di RNA possono svolgere, da soli o associati a proteine, il ruolo di elementi di regolazione in *trans*. Per la trattazione di vari tipi e classi di RNA regolatori, sia in procarioti che in eucarioti, rinviamo al capitolo 18 che è specificamente dedicato a nuovi e sorprendenti proprietà e ruoli dell'RNA.

Lecture di approfondimento

- Chang, Y.F., Imam, J.S. e Wilkinson, M.F. (2007), «The nonsense-mediated decay RNA surveillance pathway», *Annual Review of Biochemistry*, **76**, pp. 51-74.
- Holt, C.E. e Bullock, S.L. (2009), «Subcellular mRNA localization in animal cells and why it matters», *Science*, **326**, pp. 1212-1216.
- Gebauer, F. e Hentze, M.W. (2004), «Molecular mechanisms of translational control», *Nature Reviews Molecular Cell Biology*, **5**, pp. 827-385.
- Sonenberg, N. e Hinnebusch, A.G. (2009), «Regulation of translation initiation in eukaryotes: mechanisms and biological targets», *Cell*, **136**, pp. 731-745.

G

ALTRI LIVELLI DI REGOLAZIONE: REGOLAZIONI EPIGENETICHE E POST-TRADUZIONALI, RUOLI DI RNA NON CODIFICANTI (ncRNA)

Nelle sezioni precedenti abbiamo visto i vari meccanismi di regolazione dell'espressione genica che operano ai tre principali livelli del flusso di informazione genetica: la trascrizione, il processamento dell'RNA e la traduzione. In questa sezione vedremo meccanismi regolativi dell'espressione genica che operano ad altri livelli, soprattutto negli eucarioti. Alcuni di questi meccanismi operano a monte della trascrizione e altri a valle della traduzione.

In particolare, nel capitolo 17 parleremo di due tipi di modificazioni genomiche che hanno effetti epigenetici sulla regolazione dell'attività genica: (1) la metilazione delle citosine nel DNA e la relazione tra questa metilazione e l'attività genica; (2) il cosiddetto "rimodellamento della cromatina", basato sull'interferenza tra organizzazione della cromatina e trascrizione.

Nel capitolo 18 vedremo il ruolo di specie di RNA diverse da quelle implicate direttamente nella sintesi proteica. Questi RNA, di scoperta più recente, sono genericamente chiamati ncRNA (non-coding RNA) e comprendono vari tipi di RNA, tra cui microRNA, lncRNA, RNA circolari ecc. che svolgono molteplici ruoli nella regolazione di numerosi processi biologici e che suggeriscono e supportano il concetto di Mondo a RNA. Questo avrebbe caratterizzato l'origine e le prime fasi della storia della vita sulla Terra. Da questo sarebbero poi scaturiti ed evoluti gli altri componenti macromolecolari degli organismi viventi quali, soprattutto, proteine e DNA, che avrebbero assunto funzioni più specifiche e consone all'evoluzione di organismi viventi.

Infine, nel capitolo 19 parleremo delle più importanti modificazioni post-traduzionali che rappresentano l'ultimo livello di regolazione nel flusso dell'informazione genica.

17

Regolazioni epigenetiche: metilazione del DNA e rimodellamento della cromatina

17.1 Metilazione del DNA ed espressione genica

17.2 Rimodellamento della cromatina

Come spiegato nella **finestra 17.1**, il termine “eredità epigenetica”, nel contesto delle odierne Genetica e Biologia Molecolare, si riferisce alla trasmissione ereditaria di caratteri fenotipici causata da meccanismi diversi da alterazioni di sequenze nucleotidiche nel DNA. Cambiamenti di questi caratteri possono essere trasmessi dalla cellula madre alle cellule figlie attraverso la mitosi e, in alcuni casi, anche attraverso la meiosi. I due principali meccanismi di controllo epigenetico dell’attività genetica sono correlati alla metilazione del DNA e al rimodellamento della struttura della cromatina. Infatti, alla base della regolazione epigenetica c’è la trasmissione alle cellule figlie dello stato “espressivo” di un gene, cioè esprimibile/non esprimibile, che, come vedremo, è a sua volta determinato dalla struttura locale della cromatina e dalla metilazione del DNA, caratteristiche strutturali che possono entrambe essere conservate attraverso numerosi cicli di duplicazioni del genoma.

La rilevanza dei processi epigenetici per quasi tutti gli aspetti della normale crescita e sviluppo è dimostrata dall’esistenza, nell’uomo, di un gran numero di disordini multisistemici e neoplasie correlati al loro mancato funzionamento.

17.1 Metilazione del DNA ed espressione genica

Abbiamo avuto già varie occasioni di menzionare la presenza negli acidi nucleici di nucleotidi modificati: presenza di basi anomale, metilazioni delle basi canoniche e metilazioni dello zucchero. Questi nucleotidi modificati non vengono inseriti nel DNA durante la replicazione, o nell’RNA durante la trascrizione, ma sono sempre il risultato di modificazioni, da parte di enzimi proteici o ribonucleoproteici, dei normali nucleotidi già inseriti nelle molecole di DNA o RNA. Le modificazioni sono altamente specifiche, riguardando particolari siti e sequenze, e hanno diversi importanti ruoli sulla struttura e funzione delle sequenze nucleotidiche di acidi nucleici.

Per quanto riguarda in particolare la metilazione del DNA, abbiamo già visto (*vedi* cap. 7, *finestra* 7.1) che nei procarioti essa è implicata nel distinguere il filamento di nuova sintesi dal filamento parentale durante vari processi di replicazione e di riparazione del DNA. Vedremo poi (*vedi* par. 20.4) che la metilazione di alcune basi del DNA nei procarioti è implicata nel processo di modificazione/restrizione che è alla base del fenomeno della “restrizione d’ospite”.

Consideriamo ora il ruolo della metilazione del DNA negli eucarioti; tale metilazione riguarda principalmente la citosina, che viene modificata in 5-metilcitosina (**fig. 17.1**). Questa modificazione ha un importante ruolo nella regolazione

FINESTRA 17.1 EPIGENESI, EPIGENETICA E REGOLAZIONI EPIGENETICHE

I termini “epigenesi” ed “epigenetica”, pur riferendosi ambedue a processi relativi allo sviluppo degli organismi complessi, indicano fenomeni sostanzialmente diversi. Inoltre, il loro uso approssimato e ambiguo ha generato alle volte una notevole confusione. La parola **epigenesi** è stata introdotta nella seconda metà dell'Ottocento a indicare una delle due contrapposte teorie dello sviluppo. Si riferiva, infatti, a un meccanismo di sviluppo che, partendo da un qualcosa di semplice e senza forma, porta al progressivo differenziamento e sviluppo di tessuti e organi fino alla formazione di un organismo complesso. Questa visione era contrapposta alla teoria preformista, poi abbandonata, che interpretava lo sviluppo come la semplice crescita di un organismo piccolissimo ma già completamente preformato.

Il termine **epigenetica**, che vuol dire “al di sopra della genetica” od “oltre la genetica”, è stato introdotto nel 1942 da Conrad Waddington ed è stato usato nel tempo con diversi significati. Inizialmente, veniva usato alquanto genericamente per indicare lo studio delle interazioni tra

geni e ambiente nella formazione dell'organismo. L'enfasi sul ruolo dell'ambiente ha in alcuni casi indebitamente accostato l'epigenetica a una visione neo-lamarckiana della Biologia. È stato poi usato per indicare in maniera diretta i meccanismi di controllo temporale e spaziale dell'attività dei geni durante lo sviluppo di organismi complessi fino a che ha acquistato un più preciso significato riferendosi a quegli aspetti, diversi e aggiuntivi alla sequenza del DNA, che influenzano lo sviluppo di un organismo. Quest'ultima è l'accezione con cui il termine “epigenetica” è oggi più frequentemente usato, cioè a indicare casi anomali di ereditarietà in cui si osservano cambiamenti fenotipici che, non dipendendo da cambiamenti della sequenza nucleotidica del DNA, apparentemente sfuggono alle classiche leggi mendeliane.

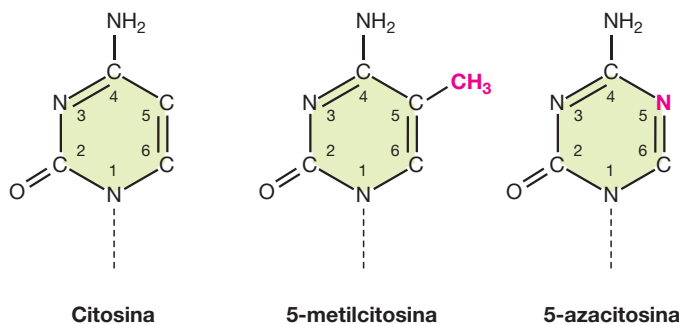
I principali meccanismi epigenetici di regolazione dell'attività genica sono stati identificati nella metilazione del DNA e nel processo di rimodellamento della cromatina, che sono trattati in questo capitolo.

dell'espressione genica, soprattutto negli eucarioti superiori: infatti, mentre nei mammiferi circa il 5% delle citosine è metilato, lo è meno dell'1% in *Drosophila* e lievito, e sarebbe assente in *C. elegans*.

Metilazione del DNA genomico nei mammiferi

Nei mammiferi sono metilate principalmente le citosine (C) seguite da G, cioè nella sequenza 5'-CG-3', che viene chiamata **CpG**, dove la “p” sta a indicare il gruppo fosfato che lega il 3' della C al 5' della G. È stato importante a questo punto lo sviluppo di tecniche per l'analisi della frequenza e distribuzione delle CpG metilate nelle sequenze genomiche *in vivo*, tecniche che sono brevemente descritte nella **finestra 17.2**. È stato osservato che non tutte le sequenze CpG sono metilate ma, per esempio, lo sono il 70-80% di tutte le sequenze CpG presenti nei genomi dei mammiferi. Come mostrato in **figura 17.2A**, se consideriamo una sequenza CpG nel contesto di un DNA a doppio filamento, possiamo osservare che essa costituisce in effetti una mini-palindrome, cioè si legge 5'-CG-3' su ambedue i filamenti complementari. Inoltre, osserviamo che le C sui due filamenti sono ambedue metilate o ambedue non metilate. Subito dopo un

Figura 17.1 Struttura della citosina e delle sue forme modificate 5-metilcitosina e 5-azacitosina.



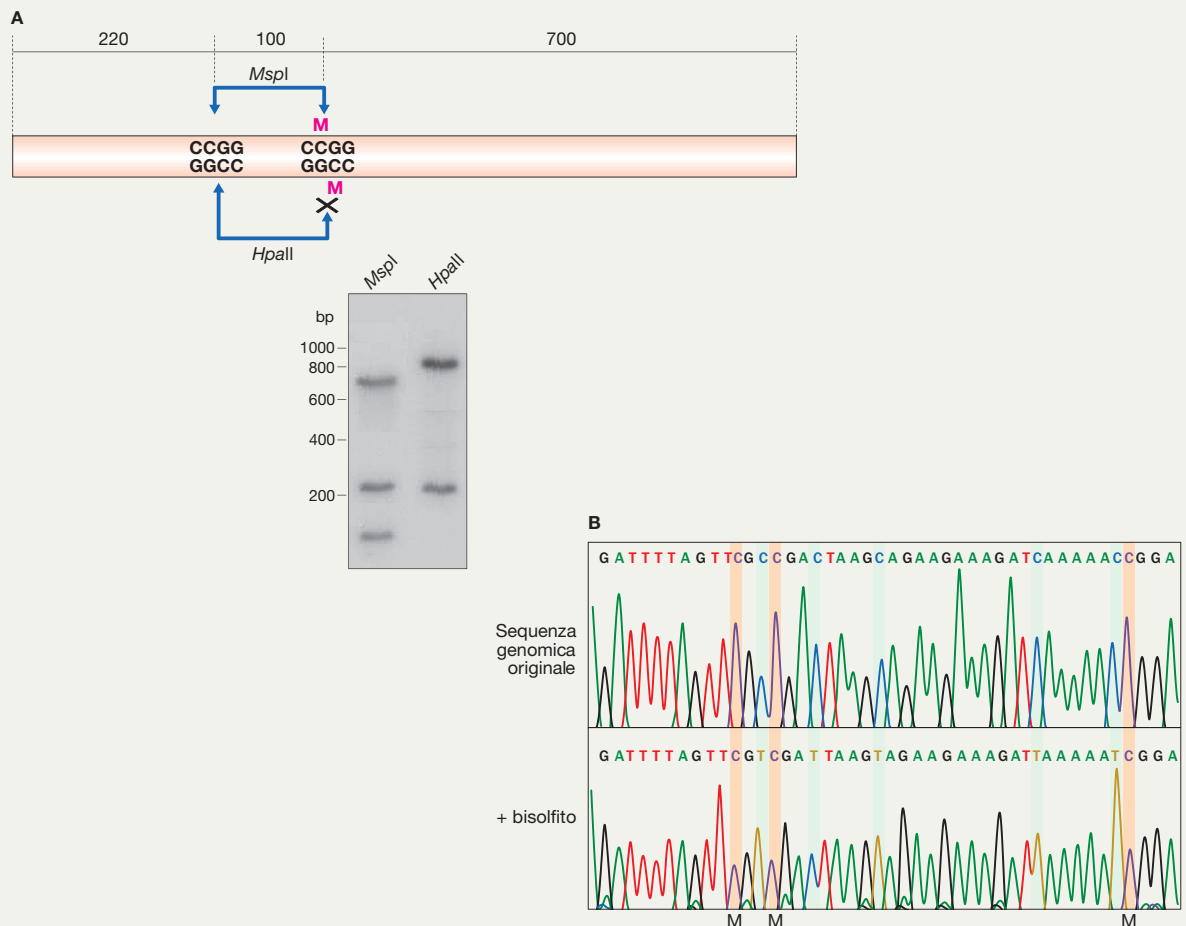
FINESTRA 17.2 IDENTIFICAZIONE DELLE CpG METILATE NEL DNA GENOMICO

Lo studio della metilazione e dei suoi effetti sulla regolazione dell'espressione genica ha trovato notevoli difficoltà sperimentali per il semplice fatto che le tecniche di isolamento, clonaggio e sequenziamento dei geni fanno sì che le metilazioni vengano perse durante i vari passaggi della procedura. Pertanto, i ricercatori hanno dovuto escogitare nuovi metodi che permettessero di analizzare lo stato metilato delle CpG direttamente nel DNA estratto dalle cellule di interesse.

Per molti anni l'unica tecnica disponibile è stata quella che sfrutta la proprietà di una coppia di enzimi di restrizione, *HpaII* e *MspI*; entrambi gli enzimi riconoscono e tagliano la sequenza palindromica 5'-CCGG-3', con la differenza che *HpaII* la taglia solo se non è metilata, mentre *MspI* la taglia sia se non metilata che se metilata sulla seconda C. La **figura** mostra come con questi enzimi di restrizione e un'analisi di Southern blot (*vedi* par. 20.2) è possibile identificare se una particolare CpG sia metilata o no. Il limite principale di questa tecnica è che essa non permette di analizzare tutte le CpG presenti in un tratto di DNA, ma solo quelle poche che si trovano nel contesto 5'-CCGG-3', bersaglio dei due enzimi di restrizione.

Successivamente, è stato introdotto un metodo che permette uno studio più approfondito e con cui è possibile analizzare il profilo di metilazione di un tratto di DNA. Questo metodo, chiamato **mappatura con bisolfito**, si basa sul fatto che il trattamento del DNA con bisolfito di sodio risulta nella modificazione della normale citosina in uracile, mentre la citosina metilata non viene modificata. Dopo il trattamento con bisolfito il DNA genomico viene amplificato mediante PCR, subclonato e infine sequenziato. Il risultato di questo sequenziamento viene confrontato con la sequenza originale. Come mostrato nella **figura** le citosine metilate e non metilate possono essere facilmente identificate perché, dopo trattamento con bisolfito, le C metilate (evidenziate in rosa) sono rimaste C, mentre le C non metilate (evidenziate in verde) sono diventate T.

L'avvento delle piattaforme di sequenziamento di nuova generazione rende oggi possibile l'applicazione della metodica del bisolfito per la determinazione della mappa di metilazione dell'intero genoma o di estese regioni di questo in un singolo ciclo di sequenziamento ed evitando la fase di clonaggio (*vedi* par. 20.8).



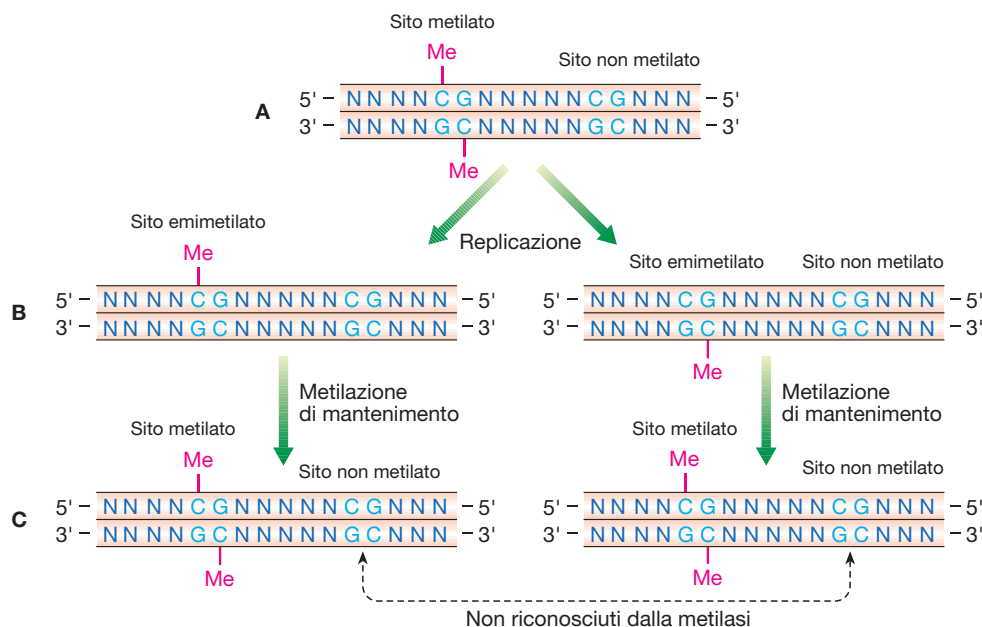


Figura 17.2 La metilasi di mantenimento perpetua lo stato di metilazione delle CpG attraverso i cicli di replicazione. (A) Tratto di DNA genomico a doppio filamento contenente una CpG metilata e una CpG non metilata. (B) Subito dopo la replicazione, ciascuna cellula figlia presenta una doppia elica emimetilata nel si-

to che era metilato nella cellula parentale e non metilata nell'altro sito. (C) La metilasi di mantenimento riconosce i siti emimetilati e metila il filamento di neosintesi, ricostituendo una CpG metilata su entrambi i filamenti; l'altro sito resta non metilato.

evento di replicazione del DNA (**fig. 17.2B**), il filamento parentale conserva la C metilata, mentre il filamento di nuova sintesi non è metilato; considerando la doppia elica avremo quindi un sito "emimetilato" che dovrà in qualche maniera diventare completamente metilato.

La metilazione delle sequenze CpG è dovuta alle **DNA metil transferasi (DNMT)**, per brevità chiamate in genere **metilasi**, enzimi che aggiungono un gruppo metilico alla posizione 5 della citosina. Esistono anche delle **demetilasi** che hanno il ruolo di rimuovere i gruppi metilici. Le metilasi si dividono in due categorie: le **metilasi de novo** di cui parleremo più avanti, che aggiungono gruppi metilici in nuove posizioni, e le **metilasi di mantenimento**. Queste ultime riconoscono i siti emimetilati e aggiungono un metile alla C del filamento non metilato, rendendo così i siti completamente metilati, mentre non riconoscono, e quindi non metilano, i siti non metilati che rimangono tali (**fig. 17.2C**). Così i siti metilati nel cromosoma parentale lo saranno anche nei cromosomi figli dopo la replicazione, mentre i siti non metilati nella doppia elica parentale resteranno tali. In altre parole, le metilasi di mantenimento garantiscono che il profilo di metilazione venga perpetuato attraverso numerosi cicli di repliche del DNA e di divisioni cellulari.

Isole CpG

È stato osservato che le sequenze CpG (metilate o non metilate) non sono distribuite uniformemente lungo la sequenza genomica, ma appaiono raggruppate in regioni lunghe circa 1-2 Kpb, chiamate **isole CpG**, che si trovano più o meno a cavallo dell'estremità 5' di molti geni, come nell'esempio mostrato in **figura 17.3A**. Le isole CpG potrebbero a prima vista venire considerate regioni ipermetilate, proprio perché contengono un gran numero di CpG che sono bersaglio della metilazione. Al contrario, è stato dimostrato che le isole CpG sono in realtà



Figura 17.3 Isole CpG: distribuzione delle CpG nel genoma umano. Sono mostrati due frammenti di DNA genomico umano contenenti: (A) il gene housekeeping *EE1A1* per il fattore di traduzione eEF1A1; (B) il gene da differenziamento *G6PC2* per la subunità ca-

talitica della glucosio-6-fosfatasi, espresso nel pancreas. Gli esoni dei due geni sono indicati in verde e numerati. La presenza di un'isola CpG nella regione del promotore e del 5' del gene *EE1A1*, ma non nel gene *G6PC2*, è evidente.

regioni di ipometilazione: vediamo perché. Va innanzitutto detto che un'isola CpG non è dovuta alla presenza di un eccesso di CpG in quella particolare regione, ma piuttosto al fatto che le altre regioni genomiche, tra un'isola e l'altra, sono particolarmente "svuotate" di CpG. Infatti, in seguito al processo di mutazione delle citosine metilate, si osserva una generale sottorappresentazione nel genoma del dinucleotide CpG, che risulta essere presente in misura pari a circa il 20% del valore atteso. Nelle isole CpG tale valore supera il 60%. La causa di questo svuotamento va cercata nel fatto che le metilcitosine presenti nel DNA possono abbastanza frequentemente mutare in altre basi, o perché la metilcitosina può andare incontro a deaminazione spontanea ed essere così convertita in timina, o perché durante la replicazione la DNA polimerasi può commettere errori quando incontra una metilcitosina.

Questa maggior mutabilità della metilcitosina risulta, nei tempi lunghi dell'evoluzione, in un progressivo impoverimento di C nel genoma, ma solo di quelle C che possono essere metilate, quindi quelle seguite da una G. Pertanto, le isole CpG rappresentano regioni genomiche protette dalla metilazione; queste sono soprattutto le regioni intorno ai promotori, che, essendo perlopiù legate a fattori e a complessi proteici implicati nella trascrizione e nella sua regolazione, sono poco accessibili alle metilasi. Più un gene è espresso nei diversi tessuti e in varie situazioni, tanto più le sue CpG intorno al promotore sono protette dalla metilazione.

In accordo con questa nozione è stato trovato che sostanzialmente tutti i geni housekeeping (cioè espressi in tutte le cellule) sono caratterizzati da isole CpG, mentre lo sono meno del 40% degli altri geni, per esempio quelli implicati nel differenziamento. La **figura 17.3A e B** mette a confronto il gene housekeeping per il fattore traduzionale eEF1A1 e il gene *G6PC2* per la subunità catalitica della glucosio-6-fosfatasi, che è espresso specificamente nel pancreas. Come si può osservare, un'isola CpG è presente nella regione 5' del gene per eEF1A1, ma non in quella del gene *G6PC2*.

Metilazione delle CpG e trascrizione

La metilazione del DNA ha un importante ruolo nella regolazione dell'attività genica a livello di trascrizione. In genere l'ipermetilazione delle sequenze CpG nella regione del promotore di un gene corrisponde a un'inibizione della trascrizione e, al contrario, un'ipometilazione correla con un'attivazione della trascrizione.

Uno degli esperimenti classici che dimostrarono il ruolo negativo della metilazione sull'espressione genica è il trattamento di colture cellulari con 5-azacitosina, un analogo della citosina che, somministrato a cellule in coltura, viene incorporato nel DNA al posto delle C. Poiché, come mostrato in **figura 17.1**, il carbonio 5 della citosina è sostituito da un azoto nella azacitosina, questa non può essere metilata, risultando nell'attivazione trascrizionale di vari geni. Un'altra osservazione che va in questa direzione è che la metilazione delle CpG è relativamente alta nel genoma di un mammifero ai primissimi stadi di sviluppo e va progressivamente diminuendo durante il differenziamento e lo sviluppo dell'embrione, in parallelo con l'attivazione dei vari geni per il differenziamento.

La metilazione del DNA può influire sulla trascrizione dei geni in due diverse modi: (1) un fattore di trascrizione può non riconoscere la sua sequenza bersaglio in cui una C sia metilata, con conseguente inibizione della trascrizione; (2) in altre situazioni, la sequenza metilata viene riconosciuta e legata da alcune proteine, note come **proteine MBD** (*Metil-CpG-Binding Domain Protein*), che a loro volta reclutano altri fattori implicati nel rimodellamento della cromatina rendendola inattiva. Questa connessione tra metilazione del DNA e struttura cromatinica è molto interessante e ci torneremo poco più avanti. Va sottolineato che solo alcune CpG sono coinvolte nella regolazione trascrizionale di un gene e che esse si trovano, in genere, nella regione del promotore, o comunque intorno al 5' del gene, indipendentemente dal fatto che vi siano poche CpG o tante fino a costituire un'isola CpG.

Va anche sottolineato che, benché la metilazione delle CpG, soprattutto nella regione intorno al promotore, sia sicuramente un possibile meccanismo di silenziamento genico, non si può genericamente concludere che la metilazione/demetilazione sia un meccanismo generale di regolazione della trascrizione. Infatti, in alcune situazioni è stato osservato che la demetilazione è necessaria, ma non sufficiente, per l'attivazione di un gene: per esempio, le CpG della regione del promotore di un gene inattivo localizzato in una regione di eterocromatina facoltativa sono solitamente metilate, ma quando il gene deve essere espresso, devono venire demetilate per permettere ai fattori di trascrizione di svolgere il loro ruolo. In altre situazioni, invece, la demetilazione appare essere una conseguenza, piuttosto che una causa, dell'attivazione di un gene.

Imprinting genetico

Come abbiamo detto precedentemente, il profilo di metilazione del genoma viene mantenuto uguale, durante il susseguirsi dei cicli di replicazione del DNA e quindi dei cicli cellulari, dalla metilasi di mantenimento (il termine "profilo" con cui traduciamo il termine inglese *pattern* sta qui a indicare il particolare assortimento delle CpG che sono metilate e quali no.) Questo fatto è responsabile del fenomeno che va sotto il nome di **imprinting genetico**, che, insieme al rimodellamento della cromatina di cui parleremo più avanti, rappresenta il principale meccanismo alla base di fenomeni epigenetici. L'imprinting genetico, che riguarda solo un numero limitato di geni dei mammiferi, risulta nell'espressione della copia genica proveniente da un genitore e non della copia proveniente dall'altro genitore, anche se le due copie hanno esattamente la stessa sequenza nucleotidica. Il motivo di questa diversa espressione sta nel fatto che le due copie

geniche possono essere diversamente metilate. Quindi la metilazione “contrassegna” la copia genica come di provenienza materna o paterna, da cui il termine “imprinting”.

Come abbiamo già detto, la metilazione delle CpG è relativamente alta nel genoma di un mammifero ai primissimi stadi di sviluppo e va progressivamente diminuendo durante il differenziamento di tessuti e organi. Solo nelle cellule germinali che si sviluppano nell’embrione, sia maschio che femmina, si ha una completa demetilazione dell’intero genoma. Successivamente, durante la spermatogenesi o l’oogenesi, le metilasi *de novo* ristabiliscono, con meccanismi non ancora chiariti, la metilazione secondo un preciso profilo, che è però un po’ diverso nei maschi e nelle femmine: mentre la maggior parte dei geni hanno metilate le stesse CpG nel maschio e nella femmina, alcuni geni presentano metilazione nel maschio e non nella femmina, o viceversa. Comunque sia, solo per i geni che presentano questo profilo di metilazione sesso-specifico, dopo l’unione dei due gameti nel nuovo embrione, la copia proveniente da un genitore sarà metilata e inattiva, mentre quella proveniente dall’altro genitore sarà non metilata e attiva. Se le due copie geniche sono identiche, il fatto che si esprima solo la copia di origine materna o solo quella di origine paterna non fa alcuna differenza. Se invece una delle due copie è un allele mutato non funzionale, ci sarà un effetto diverso se esso proviene dal padre o dalla madre.

La **figura 17.4** illustra in modo schematico l’effetto dell’imprinting genetico sul fenotipo. Nella realtà dei fatti la situazione può essere notevolmente più complessa per vari motivi: innanzitutto, gran parte delle varie decine di geni dei mammiferi che presentano imprinting si trovano raggruppate in pochi siti genomici in cui i diversi geni, alcuni metilati nel maschio e altri nella femmina, si influenzano l’un l’altro; secondariamente, la metilazione non riguarda necessariamente le CpG nella regione intorno al promotore ma anche regioni vicine ma non adiacenti al gene stesso; infine, l’effetto della metilazione è spesso indiretto passando attraverso cambiamenti strutturali della cromatina di cui parleremo più avanti. Come descritto nella **finestra 17.3**, la perdita dell’imprinting è coinvolta in un certo numero di malattie ereditarie la cui trasmissione appare trasgredire le classiche leggi della genetica.

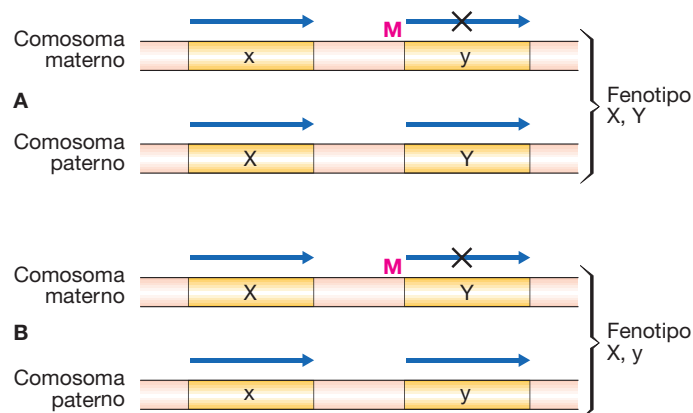


Figura 17.4 Metilazione del DNA e imprinting. In questo modello, schematico e semplificato, sono mostrati due geni allo stato eterozigote X/x e Y/y , di cui solo il secondo è sotto effetto dell'imprinting. Infatti esso si trova in un locus metilato e quindi trascrizionalmente inattivo nel cromosoma di origine materna, e non metilato e quindi trascrizionalmente attivo in quello di origine paterna. (A) e (B) mostrano la situazione in due individui che hanno la stessa composizione allelica, ma (A) presenta gli alleli dominanti

X e Y nel cromosoma paterno, mentre (B) presenta gli alleli dominanti nel cromosoma materno. Il risultato di questa distribuzione degli alleli è che, pur avendo la stessa situazione genetica, i due individui (A) e (B) presentano diverso fenotipo, rispettivamente dominante Y e recessivo y . Infatti in (B) l'allele dominante Y non si può esprimere a causa della metilazione. Invece, per il gene non sottoposto a imprinting, i due individui (A) e (B) presenteranno ambedue il fenotipo dominante X .

FINESTRA 17.3 REGOLAZIONI EPIGENETICHE E PATOLOGIA

Lo studio dei fenomeni e delle regolazioni epigenetiche è stato fin dall'inizio strettamente legato a studi di patologia umana. Non tanto per l'ovvio motivo che fenomeni di questo tipo sono alla base di alcune malattie ereditarie, ma soprattutto per il motivo opposto, cioè che proprio lo studio di malattie ereditarie che presentavano una ereditarietà anomala rispetto alle classiche leggi della Genetica ha permesso di individuare e studiare fenomeni di regolazione epigenetica.

Tra i processi epigenetici implicati nelle malattie umane ci sono casi di metilazione del DNA, modificazioni post-traduzionali degli istoni e cambiamenti strutturali dei nucleosomi, processi che, come abbiamo visto nel testo, influenzando la struttura della cromatina sono implicati nella regolazione dell'espressione genica. Vediamone alcuni esempi senza entrare troppo nei dettagli.

Una deregolazione di geni sottoposti a imprinting da metilazione delle citosine è alla base di alcune malattie del sistema nervoso, tra cui la **sindrome di Rett**, la **sindrome di Prader-Willi (PWS)** e la **sindrome di Angelman (AS)**. Queste ultime due coinvolgono la delezione dello stesso locus, chiamato locus PWS/AS, che si trova sul cromosoma 15. Si tratta di un intero dominio genomico di circa 2 Mpb che contiene numerosi geni e che è sottoposto a imprinting. In ambedue le malattie si ha una delezione dell'intero locus in uno dei due cromosomi omologhi. Il fenotipo risultante però, pur coinvolgendo sempre lo sviluppo del sistema nervoso con risvolti sia a livello fisico che comportamentale, è in parte diverso a seconda se la delezione riguarda il cromosoma di origine paterna (PWS) o quello di origine materna (AS). Infatti, il locus contiene almeno un gene sottoposto a imprinting paterno e almeno un gene sottoposto a imprinting materno. Lo stato di metilazione di questi geni bersaglio dell'imprinting, e quindi la loro attivazione, è sotto il controllo di un sito regolatore in *cis* (o *cis*-agente) chiamato ICR (*Imprinting Control*

Region) o anche DMA (*Differentially Methylated Domain*) che, a complicare la situazione, è a sua volta regolato dalla metilazione. Nel caso di delezione dell'ICR, l'imprinting è abolito e quindi i geni bersaglio si comportano ugualmente sia che si trovino sul cromosoma di origine paterna che su quello di origine materna.

Uno dei vari esempi di malattie genetiche che implicano alterazioni del rimodellamento della cromatina è la **sindrome CHARGE** (*Coloboma, Heart defect, Atresia choanae, Retarded growth and development, Genital hypoplasia, Ear anomalies/deafness*), che, come si evince dall'acronimo usato come nome, presenta un fenotipo che indica la deregolazione di un vasto numero di geni. È stato infatti recentemente trovato che pazienti per i quali è stata diagnosticata la sindrome CHARGE sono caratterizzati da aploinsufficienza di CHD7, un rimodellatore della cromatina ATP-dipendente, omologo a SWI/SNF di lievito, espresso prevalentemente in cervello, rene e muscoli scheletrici. Come in genere i membri della superfamiglia SWI/SNF, anche CHD7 influenza l'espressione di numerosi geni, risultando così nel complesso fenotipo che caratterizza i pazienti di sindrome CHARGE.

Il fatto che, tra le caratteristiche riscontrate nelle cellule tumorali, ci sia un generale cambiamento del quadro epigenetico, ha suggerito che alterazioni dei processi epigenetici possono essere alla base dei cambiamenti di attività genica che accompagnano la trasformazione maligna delle cellule. Infatti, è stato poi dimostrato che durante le fasi di iniziazione e progressione delle cellule tumorali si ha una notevole riprogrammazione dei vari componenti del macchinario epigenetico, tra cui la metilazione del DNA, le modificazioni istoniche e il riposizionamento dei nucleosomi.

Quindi il cancro, tradizionalmente considerato una malattia esclusivamente genetica, è oggi considerato coinvolgere, in parallelo ad alterazioni genetiche, anche anomalie epigenetiche.

17.2 Rimodellamento della cromatina

Abbiamo visto nel paragrafo 4.8 come negli eucarioti il DNA genomico sia assemblato con le proteine istoniche a formare i nucleosomi, che a loro volta sono impacchettati in strutture superiori: la fibra da 30 nm, il solenoide e le anse cromosomiche. Questo tipo di organizzazione, oltre a fornire una soluzione al problema dell'impacchettamento del genoma, offre opportunità di nuovi meccanismi di regolazione genica a livelli superiori.

Mentre nei procarioti la regolazione della trascrizione è data sostanzialmente dall'interazione tra gli elementi del promotore e fattori trascrizionali negativi e positivi, negli eucarioti questa interazione deve fare i conti con la struttura cromatinica della regione che può o meno nascondere gli elementi del promotore rendendoli inaccessibili ai fattori trascrizionali (*vedi* cap. 11). Poiché l'interazione tra ottamero istonico e DNA è notevolmente stabile, i nucleosomi non possono essere spostati direttamente dai fattori di trascrizione. È per questo che l'attiva-

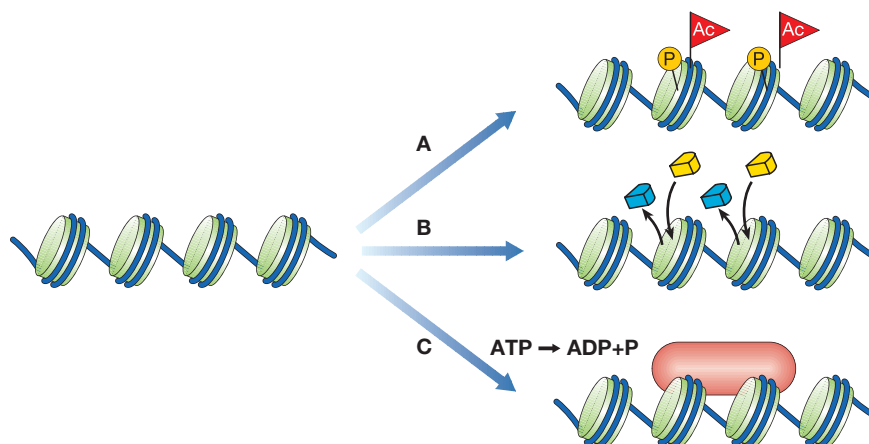


Figura 17.5 Tre tipi di processi sono implicati nel rimodellamento della cromatina. (A) Modificazioni post-traduzionali degli istoni quali acetilazioni, fosforilazioni, metilazioni e altre. (B) Sostituzione di varianti istoniche. (C) Azione di “complessi di rimodellamento della cromatina” o “rimodellatori” ATP-dipendenti.

zione di un gene richiede che, preventivamente, la struttura cromatinica subisca le opportune modificazioni in modo che gli elementi del promotore diventino accessibili per la successiva attivazione da parte dei fattori trascrizionali. In altre parole, il rimodellamento della cromatina rende un gene “attivabile”; l’effettiva attivazione dipende dal successivo intervento dei fattori. I meccanismi regolativi basati su cambiamenti della conformazione della cromatina vanno sotto il nome di **rimodellamento della cromatina**. Negli eucarioti superiori, i processi di rimodellamento della cromatina sono strettamente regolati e avvengono con un preciso ordine temporale e spaziale durante lo sviluppo. Questi cambiamenti sono generalmente ereditabili, cioè trasmessi dalla cellula madre alle cellule figlie, contribuiscono a stabilire e mantenere gli specifici programmi di espressione genica che andranno a determinare l’identità di tessuti e organi. Va anche detto che non solo la trascrizione, ma anche altri processi che avvengono a livello di genoma, quali replicazione, riparazione e ricombinazione del DNA, sono influenzati dall’organizzazione della cromatina e possono, quindi, essere regolati dal rimodellamento.

Questi cambiamenti strutturali della cromatina con effetto sulla regolazione genica sono dovuti a vari processi, spesso sovrapposti e collegati tra loro che, come mostrato in **figura 17.5**, possiamo schematicamente distinguere in tre tipi: (1) modificazioni post-traduzionali degli istoni; (2) intervento di specifici fattori o complessi proteici che prendono genericamente il nome di “complessi di rimodellamento della cromatina”; (3) sostituzione di varianti istoniche.

Modificazioni post-traduzionali degli istoni

Abbiamo già detto nel paragrafo 4.8 che gli istoni possono subire una serie di modificazioni post-traduzionali (acetilazioni, fosforilazioni, metilazioni e altre) di residui amminoacidici concentrati soprattutto alla loro estremità N-terminale. Queste modificazioni hanno un effetto sulla stabilità del nucleosoma: per esempio, le acetilazioni e le fosforilazioni, diminuendo il carattere basico degli istoni, rendono in genere meno stabile l’interazione tra l’ottamero istonico e il DNA carico negativamente, favorendo così lo spostamento del nucleosoma stesso. Le modificazioni degli istoni non sono spontanee, ma richiedono l’intervento di vari tipi di enzimi tra cui le **istone acetil transferasi (HAT)**, le **istone deacetilasi (HDAC)** e le **istone metil transferasi (HMT)**. Le modificazioni istoniche

verranno trattate più sistematicamente nel paragrafo 19.6, in relazione alle varie modificazioni post-traduzionali che si verificano su moltissime proteine.

Un'acetilazione transitoria degli istoni (H3 e H4) viene riscontrata durante la fase S del ciclo cellulare in concomitanza con la replicazione del DNA. Al di fuori di questa fase, l'acetilazione degli istoni correla con lo stato trascrizionale dei geni: istoni acetilati sono caratteristici di geni allo stato attivo, mentre istoni deacetilati sono presenti soprattutto nell'eterocromatina inattiva. Al contrario dell'acetilazione, la metilazione degli istoni è, in genere, associata all'inattività del gene, come lo è la metilazione delle CpG nel DNA stesso. Dati recenti hanno anche messo in evidenza, in alcune situazioni, una relazione tra metilazione degli istoni e metilazione delle CpG che si trovano nella regione del promotore; è stato suggerito che la metilazione degli istoni possa rappresentare un segnale per il reclutamento della DNA metilasi sulla cromatina nella regione genomica da silenziare. Questi effetti possono essere limitati a un singolo gene o a regioni cromatiniche più o meno estese o, talvolta, anche a un intero cromosoma, come è il caso di uno dei due cromosomi X nelle cellule delle femmine dei mammiferi.

Le HAT, HDAC e HMT non presentano specificità di legame a particolari siti cromosomici; la specificità viene data da particolari fattori attivatori che legano siti particolari del DNA mediante specifici domini di riconoscimento. Il promotore a valle è, invece, completamente inaccessibile per la polimerasi e il complesso di pre-inizio. L'attivatore legato a monte recluta sulla regione una istone-acetil transferasi, che acetila le code degli istoni contigui e promuove l'apertura di tutta la regione del promotore, dove ora si potrà formare il complesso di pre-inizio. Come già detto nel paragrafo 4.8, le code acetilate sono il bersaglio di particolari proteine che contengono il bromodominio, una struttura molto conservata costituita da un fascio di 4 α -eliche. Questo dominio è presente in uno dei TAFs, per cui TFIID e il complesso di trascrizione si legano con più stabilità ai nucleosomi acetilati.

Complessi di rimodellamento (rimodellatori) ATP-dipendenti

L'acetilazione degli istoni è probabilmente necessaria per decondensare la cromatina nelle regioni dei promotori, ma non è sufficiente per smontare la struttura dell'ottamero istonico nei nucleosomi. D'altra parte, come già anticipato nel paragrafo 11.7, senza la rimozione o lo spostamento dei nucleosomi, la trascrizione non può avvenire, come suggerito dal fatto che i promotori e gli enhancer dei geni attivi si trovano solitamente in regioni prive di nucleosomi. Per questo spostamento dei nucleosomi dalle regioni dei promotori è necessario l'intervento di particolari complessi proteici chiamati **complessi di rimodellamento o rimodellatori**. Questi rimodellatori, che per la loro funzione hanno bisogno di energia che è fornita dall'idrolisi di ATP, destabilizzano i nucleosomi, li scalzano o li riposizionano. In ogni caso, il risultato è la liberazione del tratto di DNA corrispondente alla regione del promotore che viene reso accessibile ai fattori trascrizionali. Poiché i complessi di rimodellamento non hanno specificità per particolari siti bersaglio, è necessario l'intervento di altri fattori specifici. Questi riconoscono e legano la regione del gene bersaglio da dove reclutano il complesso di rimodellamento. Dopo che questo ha avviato la sua attività ATP-dipendente di modificazione della cromatina, il fattore può anche essere rilasciato. Un modello dell'attivazione di un gene da parte di un rimodellatore è schematizzato in **figura 17.6**.

Ci sono due tipi principali di complessi di rimodellamento. Il primo tipo è chiamato **complesso SWI/SNF**, composto da almeno 11 subunità tra cui le proteine SWI e SNF. Queste due proteine sono state inizialmente scoperte nel lievito, dove hanno l'effetto di contrastare la repressione della trascrizione. Esse formano un complesso che ha attività ATPasica. Proteine omologhe a SWI e SNF sono

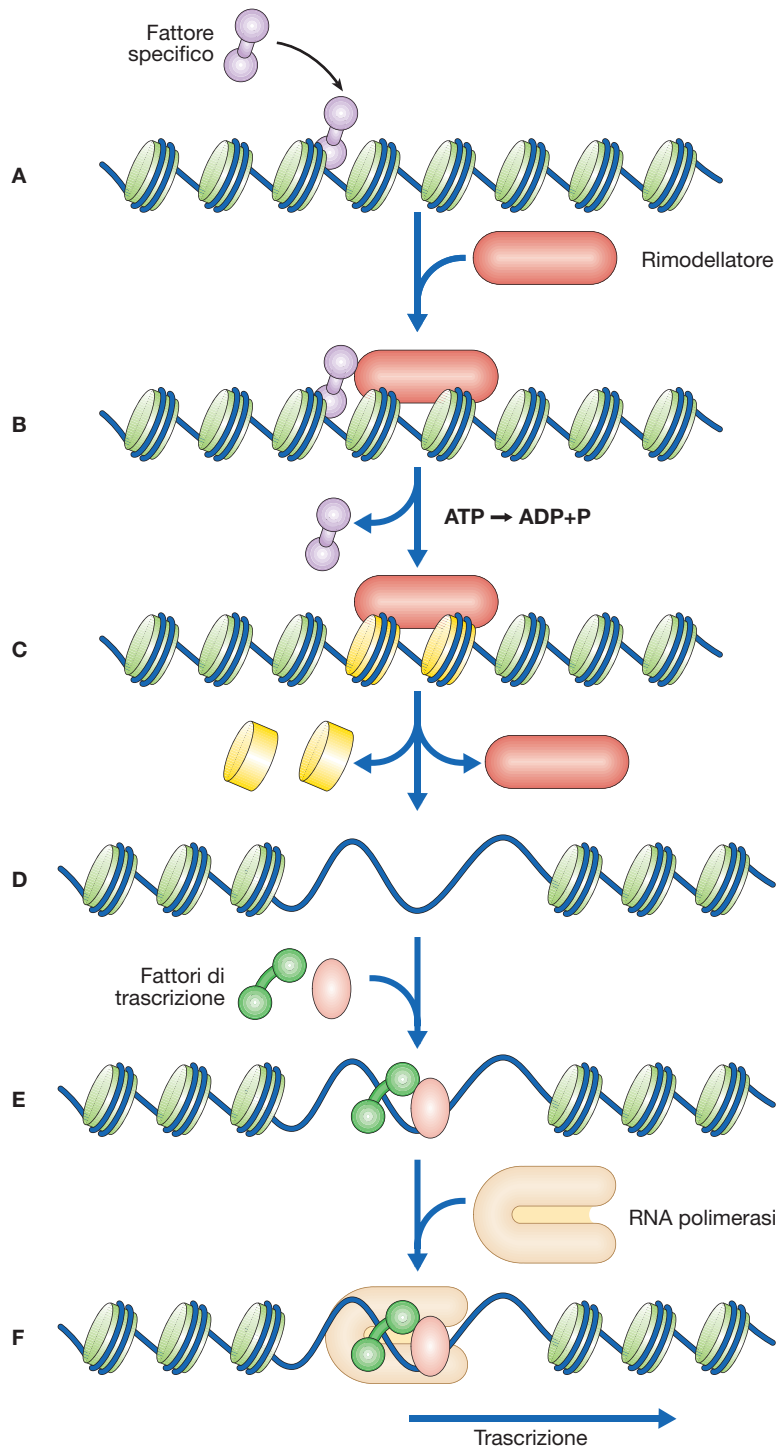


Figura 17.6 Modello schematico dell'attivazione di un gene da parte di un "rimodellatore" ATP-dipendente. (A) Un fattore specifico riconosce e lega una sequenza di DNA esposta nella regione cromatinica adiacente al gene che deve essere attivato. (B) Il fattore recluta il complesso di rimodellamento. (C) Il fattore viene rilasciato mentre il rimodellatore, utilizzando l'energia rilasciata dall'idrolisi di ATP, induce cambiamenti strutturali della cromatina che risultano in un indebolimento delle interazioni istoni-DNA. (D) I nucleosomi vengono rilasciati (o spostati sul DNA) in modo da rendere accessibile la regione del promotore. (E) I fattori di trascrizione possono ora riconoscere e assemblarsi sul promotore. (F) I fattori di trascrizione reclutano la RNA polimerasi, che può iniziare la trascrizione stessa.

state poi scoperte in tutti gli eucarioti. La seconda famiglia di rimodellatori è chiamata **ISWI** e ne comprende vari sottotipi (NURF: *NUcleosome Remodeling Factor*; CHRAC: *CHRomatin Accessibility Complex*; ACF: *ATP-dependent Chromatin-remodeling Factor*). Mentre il complesso SWI/SNF rompe il contatto DNA-istoni e provoca il rilascio dei nucleosomi dal DNA, i rimodellatori della famiglia ISWI li fanno scivolare lungo la molecola del DNA riposizionandoli in modo da lasciare libera la regione del promotore che deve interagire con i fattori di trascrizione.

Sostituzione di varianti istoniche

Abbiamo già detto nel paragrafo 4.8 che oltre ai 5 tipi fondamentali di istoni che costituiscono normalmente i nucleosomi, esistono anche alcune varianti istoniche, soprattutto varianti di H2A e H3, che influenzano l'interazione dell'ottamero istonico con il DNA. Mentre gli istoni normali vengono incorporati nei nucleosomi al momento della replicazione, l'incorporazione delle varianti istoniche sembra essere indipendente dalla replicazione e implica quindi la sostituzione di un istone in un nucleosoma già formato (**fig. 17.5B**). Studi recenti hanno fornito alcune informazioni sui complessi proteici necessari per assemblare e rimpiazzare nucleosomi alternativi. Questi complessi, implicati nella sostituzione di varianti degli istoni H2A e H3, operano specificamente su particolari siti del genoma e in particolari condizioni, per esempio specifici danni al DNA. Le differenze tra istoni varianti e canonici possono avere profonde conseguenze, o per veicolare gli istoni ai loro siti di assemblaggio o per la loro funzione dopo incorporazione nella cromatina.

Come già accennato, studi recenti hanno anche individuato connessioni tra assemblaggio di nucleosomi varianti, rimodellamento della cromatina e modificazioni post-traduzionale degli istoni. Prese nell'insieme, queste osservazioni indicano che l'architettura cromosomica può essere altamente dinamica al livello più fondamentale, con conseguenze epigenetiche.

Lettere di approfondimento

Giman, T.M. e Muegge, K. (2010), «DNA methylation in early development», *Molecular Reproduction and Development*, **77**, pp. 105-113.

Ho, L. e Crabtree, G.R. (2010), «Chromatin remodelling during development», *Nature*, **463**, pp. 474-484.

Suzuki, M.M. e Bird, A. (2008), «DNA methylation landscapes: provocative insights from epigenomics», *Nature Reviews Genetics*, **9**, pp. 465-476.

Wood, A.J. e Oakey, R.J. (2006), «Genomic imprinting in mammals: emerging themes and established theories», *PLoS Genetics*, **2**, e147.

18

Ruoli regolativi dei ncRNA e l'ipotesi del mondo a RNA

18.1 RNA regolatori nei procarioti

18.2 RNA regolatori negli eucarioti

18.3 Il mondo a RNA e l'origine della vita sulla Terra

Negli anni cinquanta e sessanta del secolo scorso, erano state individuate tre classi di RNA (mRNA, tRNA e rRNA) implicati nel processo di sintesi proteica. In anni più recenti svariate tecnologie, in particolare quelle legate al sequenziamento di tutti gli RNA presenti in una cellula (*vedi* cap. 20), hanno permesso di stabilire che almeno due terzi del DNA genomico è trascritto in molecole di RNA, ma solo una parte minoritaria di questi trascritti (meno del 2%) corrisponde a mRNA codificanti per proteine. Tutti gli altri trascritti (inclusi i tRNA e gli rRNA) sono genericamente chiamati **ncRNA** (non-coding RNA o RNA non codificanti). Tuttavia, mentre le funzioni dei tRNA e degli rRNA sono note da tempo e sono state ampiamente descritte nei capitoli precedenti, le funzioni regolative della maggior parte dei ncRNA sono ancora in gran parte da chiarire, soprattutto negli organismi eucariotici.

Alcuni di questi ncRNA che, da soli o associati a proteine in complessi ribonucleoprotici, svolgono attività regolative e catalitiche e chiamati genericamente “**ribozimi**” li abbiamo già incontrati in capitoli precedenti; nel paragrafo 12.2, abbiamo descritto l'endonucleasi RNasi P che genera l'estremità 5' dei tRNA maturi. L'RNasi P è una ribonucleoproteina la cui componente primaria è un RNA (RNA M1) che esercita l'attività catalitica; vi è poi una componente proteica di supporto dell'attività dell'RNA M1. Un altro interessante e importante esempio di ribozima che abbiamo già descritto (*vedi* par. 14.1) è rappresentato dall'rRNA 23S. Questo, oltre al ruolo strutturale dell'organizzazione degli rRNA e delle r-proteine nella struttura del ribosoma, costituisce i centri attivi di decodificazione e della peptidil transferasi, svolgendo quindi un'attività catalitica. Nel paragrafo 12.3, abbiamo descritto anche struttura e funzione dei ribozimi a “**testa di martello**”. Questi ultimi sono costituiti da piccole molecole di RNA, in grado di effettuare un taglio endonucleolitico in *cis*, e il cui ruolo funzionale consiste nel processamento dei piccoli genomi a RNA di alcuni viroidi patogeni delle piante.

Abbiamo anche già descritto altri ncRNA dotati di svariati ruoli catalitici e in particolare: l'RNA **TERC** implicato, come spiegato nel paragrafo 6.6, nel mantenimento e allungamento dei telomeri cromosomici; gli **snoRNA**, coinvolti nelle modificazioni chimiche (pseudouridilazione e 2'-O metilazione delle basi) dell'rRNA eucariotico (*vedi* par. 12.4); gli **snRNA**, componenti dello spliceosoma (*vedi* par. 13.1) e, quindi, coinvolti nella rimozione degli introni di tipo I e di tipo II; gli **RNA guida** implicati nell'editing inserzionale degli mRNA (*vedi* par. 13.2).

Non parleremo oltre di questi ncRNA avendoli già trattati nei capitoli precedenti. Descriveremo qui, invece, alcuni dei numerosi altri tipi di ncRNA che sono stati scoperti nel corso degli anni e che sono per lo più implicati, da soli o associati a proteine, nella regolazione della stabilità e traduzione degli mRNA, ma anche in processi regolativi ad altri livelli.

18.1 RNA regolatori nei procarioti

Per quanto riguarda la regolazione dell'espressione genica, sembrava assodato che i meccanismi regolativi che operano a livello di trascrizione, splicing, traduzione e stabilità dei trascritti, fossero in genere basati su interazioni tra elementi in *cis*, cioè particolari sequenze nucleotidiche che si trovano nel DNA o nell'mRNA bersaglio, ed elementi in *trans*, rappresentati da fattori proteici. Questa nozione è stata confutata o estesa da recenti scoperte che mostrano come il ruolo di elementi in *trans* può, in molti casi, essere svolto da molecole di RNA anziché da fattori proteici. Ad esempio, nei procarioti, sono stati osservati numerosi piccoli RNA, inclusi nella vasta categoria dei "piccoli RNA non codificanti" (sncRNA), che riconoscono per appaiamento di basi complementari i loro mRNA bersaglio regolandone così in *trans* la traduzione o la stabilità. In alcuni casi, invece, si tratta di sequenze di RNA che agiscono in *cis* sulla molecola stessa che li porta. Descriveremo due casi come esempi di meccanismi d'azione di sncRNA nei procarioti: gli sRNA che agiscono in *trans* e un "riboswitch" che agisce in *cis*.

sRNA

Nei procarioti l'espressione di alcuni geni o gruppi di geni è regolata da sncRNA, genericamente chiamati **sRNA**, implicati nell'adattamento della cellula a cambiamenti delle condizioni ambientali. In *E. coli* sono stati descritti più di 50 di questi sRNA regolatori. In molti casi il meccanismo di funzionamento non è stato ancora chiarito, ma sicuramente alcuni di essi agiscono mediante appaiamento di basi con gli mRNA bersaglio risultando, a seconda dei casi, in un'attivazione o in una repressione della traduzione oppure nella destabilizzazione dell'mRNA stesso. Come esempio di questo tipo di regolazione descriviamo il caso dell'RNA *oxyS* di *E. coli*. In **figura 18.1A** è schematizzata (e molto semplificata) la cascata regolativa di cui *oxyS* fa parte. Quando una cellula batterica è sottoposta a stress ossidativo, essa risponde attivando una serie di geni della difesa antiossidativa. Ad esempio, l'aggiunta di acqua ossigenata al mezzo di coltura attiva OxyR, un transattivatore trascrizionale costitutivamente presente nella cellula. OxyR, attivato dall'acqua ossigenata, induce l'espressione di alcuni geni, tra cui il piccolo RNA regolatore *oxyS* che, a sua volta, va a regolare la traduzione di una decina di mRNA bersaglio, attivandola in alcuni casi e reprimendola in altri. Tra questi ultimi casi c'è *flhA*, che è, a sua volta, un attivatore dell'operone *hly*. Come agisce l'RNA *oxyS* sul suo mRNA bersaglio? Come illustrato in **figura 18.1B**, l'RNA regolatore *oxyS* di 109 nt presenta tre strutture a forcina con stelo e ansa. La prima e la terza di queste anse contengono sequenze complementari a regioni dell'mRNA bersaglio *flhA*. In particolare, l'ansa che si trova nella regione 3' di *oxyS* è complementare a una sequenza subito a monte del codone AUG di inizio e, in parte, sovrapposta con la sequenza di Shine-Dalgarno. Questa interazione impedisce il legame della subunità piccola del ribosoma e quindi inibisce la traduzione. L'appaiamento di basi tra l'sRNA e l'mRNA bersaglio è essenziale per la specificità dell'interazione, nella quale però è coinvolto anche un componente proteico, la proteina Hfq, che contribuisce allo svolgimento delle strutture secondarie degli RNA facilitando l'interazione tra mRNA e sRNA.

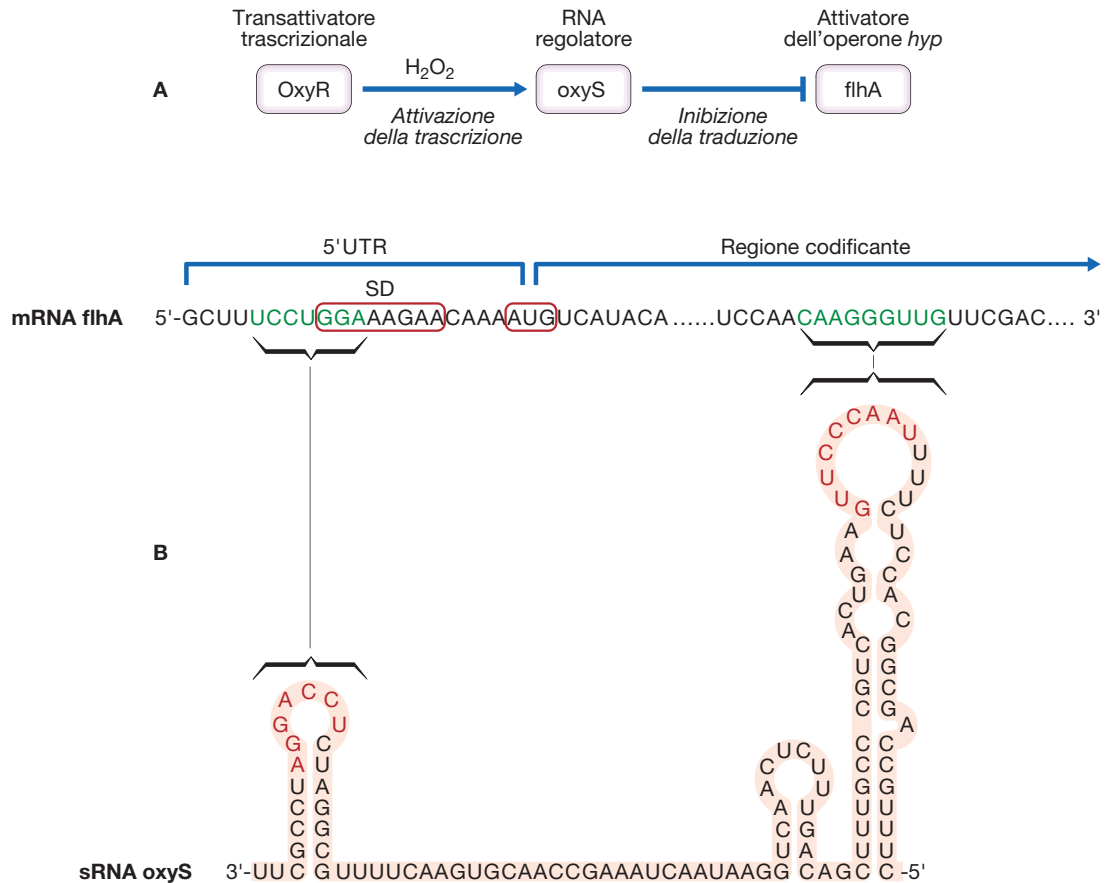


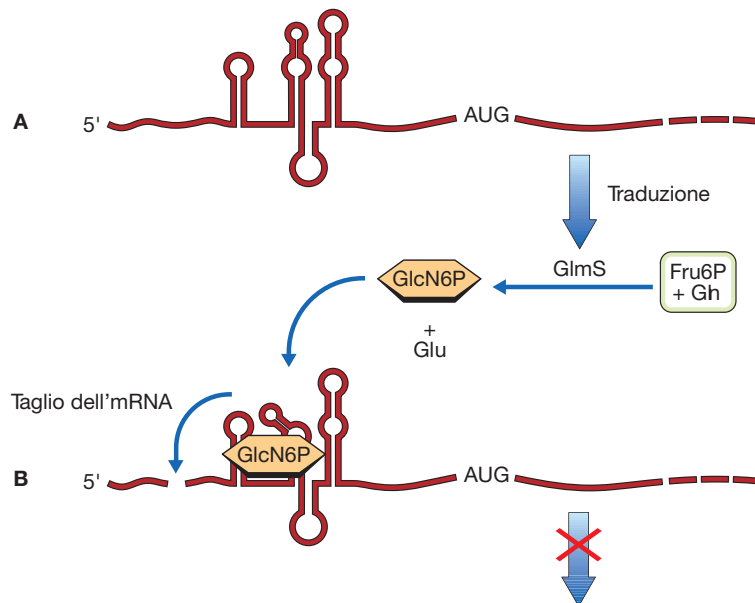
Figura 18.1 Il piccolo sRNA *oxyS* di *Escherichia coli* reprime la traduzione dell'mRNA *flhA*. (A) Lo stress ossidativo causato dall'aggiunta di H₂O₂ al mezzo di coltura attiva il transattivatore trascrizionale OxyR, che promuove la trascrizione di alcuni geni, tra cui l'sRNA *oxyS*; questo è un regolatore della traduzione di vari

mRNA bersaglio, tra cui l'mRNA *flhA*. (B) L'RNA *oxyS* riconosce l'mRNA bersaglio *flhA* mediante appaiamento di basi complementari di due regioni. Una di queste si appaia a una sequenza a monte del codone AUG e parzialmente sovrapposta alla sequenza di Shine-Dalgarno (SD), reprimendo così la traduzione.

Riboswitch

I **riboswitch** sono delle sequenze di RNA che subiscono cambiamenti conformazionali in conseguenza al legame di piccole molecole regolatrici. Essi sono alla base della regolazione di numerosi geni batterici, in genere implicati nel metabolismo di amminoacidi, nucleotidi e altre piccole molecole. I riboswitch sono parte degli stessi trascritti che essi regolano e agiscono in *cis* controllando l'espressione delle adiacenti sequenze codificanti, diversamente da quanto descritto sopra per gli sRNA che controllano in *trans* l'attività di mRNA bersaglio. La **figura 18.2** illustra l'esempio del gene *glmS* che codifica per la glucosammina-6-fosfato sintasi (GlmS), un enzima che presiede alla sintesi di glucosammina-6-fosfato (GlcN6P) a partire da fruttosio-6-fosfato e glutammina. L'mRNA per GlmS contiene nella sua 5'UTR una regione con particolare struttura secondaria che, di per sé, non ha effetti sulla stabilità e traduzione dell'mRNA stesso. Se però questa struttura viene legata da una molecola di GlcN6P, essa diventa un ribozima con attività endonucleolitica sulla stessa molecola di mRNA. Il taglio endonucleolitico rimuove la prima parte della 5'UTR dell'mRNA, che viene successivamente degradato, risultando così in una riduzione dell'espressione della proteina GlmS. Si tratta quindi di un meccanismo che garantisce un'equilibrata produzione di GlcN6P

Figura 18.2 Regolazione autogena della sintesi della GlmS mediante un riboswitch. La 5'UTR dell'mRNA che codifica per la GlmS ha la struttura di un ribozima che, se attivato, taglia lo stesso RNA che lo porta. Il ribozima viene attivato dal legame di una molecola di GlcN6P, che è il prodotto metabolico dell'enzima GlmS codificato dall'mRNA stesso.



mediante una regolazione feedback: infatti, l'accumulo di un eccesso di GlcN6P attiva il ribozima che taglia l'mRNA per la GlmS, bloccandone l'ulteriore traduzione. Riboswitch sono stati descritti soprattutto nei batteri, ma ne sono stati trovati anche in funghi e piante.

18.2 RNA regolatori negli eucarioti

Negli eucarioti, il sequenziamento del trascrittoma, insieme ad altre tecnologie, ha consentito di scoprire una grande varietà di ncRNA di lunghezza molto variabile: dai 21-23 nt dei **microRNA (miRNA)** alle migliaia di nt per la classe dei cosiddetti **long non coding RNA (lncRNA)**. Inoltre, in numerose regioni genomiche sono trascritti entrambi i filamenti, quindi nelle opposte direzioni e, in alcuni casi, questi trascritti si sovrappongono come sequenze antisense agli stessi mRNA.

MicroRNA

I microRNA (miRNA) maturi degli eucarioti sono piccoli RNA di 21-23 nt che derivano, come vedremo poco più avanti, dal processamento di molecole precursori di RNA molto più grandi e, interagendo in *trans* con molecole di mRNA bersaglio, ne influenzano la stabilità o traducibilità. Il primo microRNA è stato identificato nel 1993 da R.C. Lee, R.L. Feinbaum e V. Ambros in *C. elegans*, dove l'espressione temporale del gene *lin-14*, un gene coinvolto nello sviluppo larvale, è regolata negativamente a livello post-trascrizionale dal prodotto del gene *lin-4*, che è una piccola molecola di RNA di 22 nucleotidi. L'osservazione iniziale è illustrata in **figura 18.3A**, dove si vede che la proteina Lin14 (prodotta dal gene *lin-14*) è presente durante il primo stadio larvale, ma sparisce rapidamente negli stadi successivi in concomitanza con uno speculare aumento dell'RNA lin4. Questo è una breve molecola di RNA chiamata appunto microRNA (miRNA) che ha una regione complementare a una sequenza di 10 nt ripetuta sette volte nella 3'UTR dell'mRNA bersaglio *lin-14* (**fig. 18.3C**). L'interazione per appaiamento di basi tra miRNA *lin-4* e mRNA risulta in una riduzione

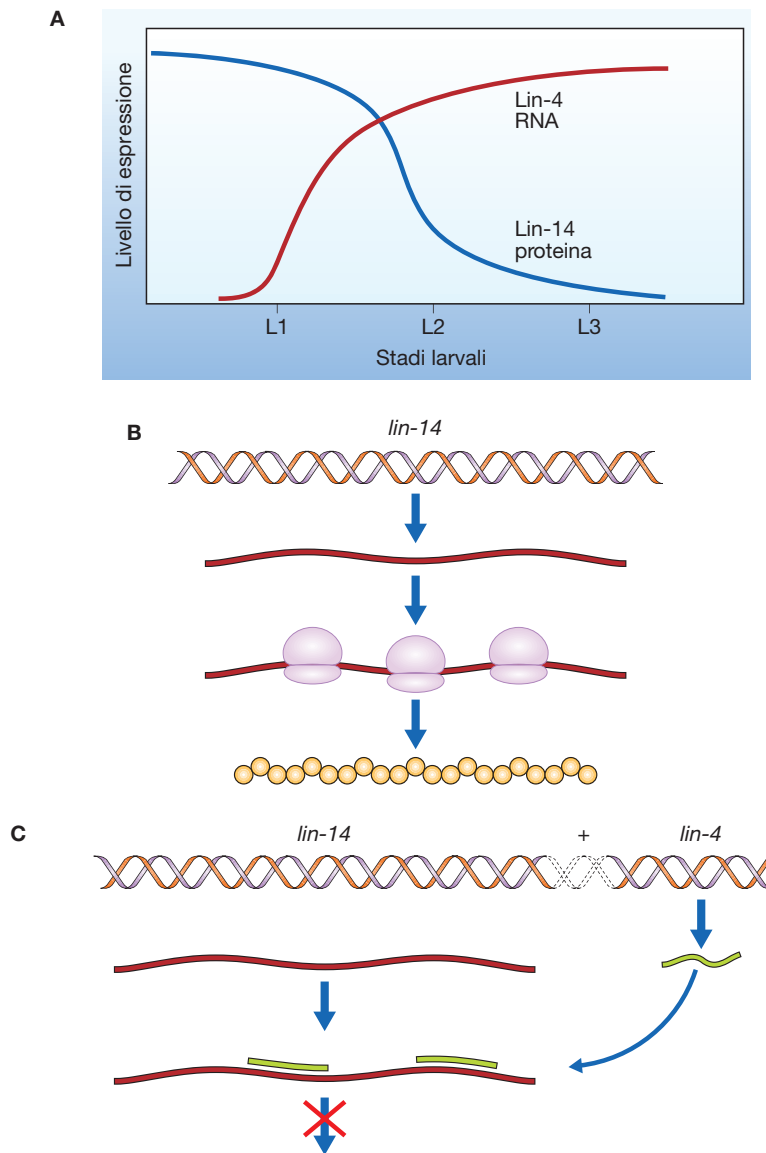
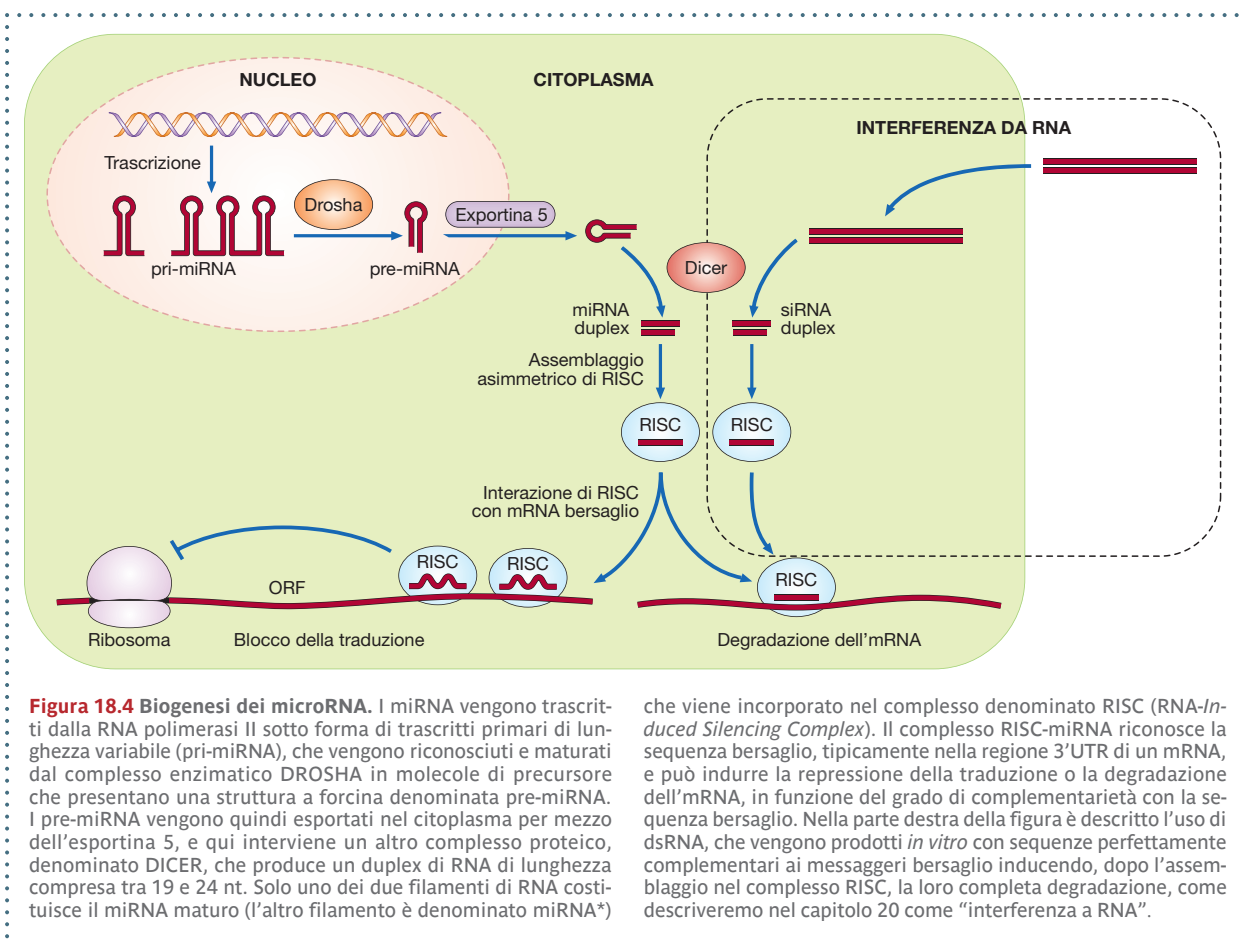


Figura 18.3 Il microRNA *Lin-4* controlla l'espressione del gene *lin-14* durante lo sviluppo larvale di *C. elegans*. (A) Curve di espressione durante lo sviluppo larvale dei geni *lin-14* e *lin-4*. Il gene *lin-14* è espresso al massimo livello durante lo stadio larvale 1 per poi decrescere rapidamente in concomitanza con l'aumento dell'espressione del gene *lin-4*. (B) Durante lo stadio larvale 1 il gene *lin-14* è attivamente trascritto e il corrispondente mRNA attivamente tradotto nel prodotto proteico Lin-14. (C) Durante lo stadio larvale 2 e seguenti, la trascrizione del gene *lin-4* porta alla sintesi di un RNA che interagisce per appaiamento di basi con l'mRNA di *lin-14* e ne blocca la traduzione.

della produzione della proteina. Successivamente, sono stati trovati numerosi altri microRNA in *C. elegans*, ma anche in altri eucarioti, animali e vegetali. Nell'uomo ne sono stati trovati centinaia e se ne scoprono ancora di nuovi. Questi RNA stanno attirando molto l'attenzione dei ricercatori per il ruolo che svolgono nella regolazione dell'espressione genica, in particolare nel controllo della stabilità e della traduzione di specifici mRNA in importanti processi quali la proliferazione e il differenziamento cellulare, l'apoptosi e lo sviluppo. Nonostante l'interesse e gli sforzi, ci sono ancora molti aspetti non chiariti sui ruoli, sui meccanismi di funzionamento e su quali siano i bersagli dei miRNA.

I miRNA vengono generati da precursori più grandi, denominati miRNA primari (pri-miRNA), che possono essere espressi da geni specifici non codificanti che contengono uno o più miRNA nelle regioni esoniche o introniche, oppure dalle regioni introniche di geni codificanti proteine. I pri-miRNA sono trascritti dalla RNA polimerasi II, maturati dai complessi DROSHA e DICER, rispettivamente nel nucleo e nel citoplasma, per poi essere assemblati nel complesso RISC, come vedremo poco più avanti. In breve, i trascritti primari prodotti dalla RNA



polimerasi II subiscono i processi di capping, splicing e poliadenilazione (vedi capp. 13 e 14). Sono poi sottoposti a un processo di maturazione sia a livello nucleare, denominato **cropping**, che a livello citoplasmatico, denominato **dicing** (fig. 18.4). Questi processi vengono operati rispettivamente dai complessi DROSHA e DICER, entrambi appartenenti alla classe delle RNasi III. Il **cropping** rimuove tratti extra di sequenza alle estremità 5' e 3' del pri-miRNA e genera una struttura a forcina (*stem-loop*) di lunghezza pari a 60-100 nt, denominata pre-miRNA e caratterizzata da una sporgenza costituita da una coppia di basi. Il pre-miRNA viene successivamente traslocato nel citoplasma dal sistema esportina-5/Ran-GTP attraverso il poro nucleare (vedi fig. 13.26). Nel **dicing** viene formato il miRNA duplex attraverso la rimozione del loop del pre-miRNA. Successivamente, viene selezionato uno dei due filamenti complementari del miRNA duplex, che costituisce il miRNA attivo e che viene incorporato nel complesso RISC, che potrà interagire con l'mRNA target.

Come anticipato, i microRNA maturi sono brevi RNA a singolo filamento della lunghezza di 22 (± 1) nt che si legano agli mRNA bersaglio per appaiamento di basi per lo più nella regione 3'UTR, ma non solo. Generalmente più miRNA diversi, o più copie dello stesso miRNA, possono legare un mRNA bersaglio contribuendo, nell'insieme, al controllo della produzione della proteina a diversi livelli. In effetti, i miRNA non funzionano da soli, ma solo se assemblati in un complesso effetore ribonucleoproteico chiamato miRNP (*miRNA-containing ribonucleo-protein particle*) o nel molto simile RISC (*RNA-Induced Silencing Complex*), che oltre al miRNA comprende alcune proteine.

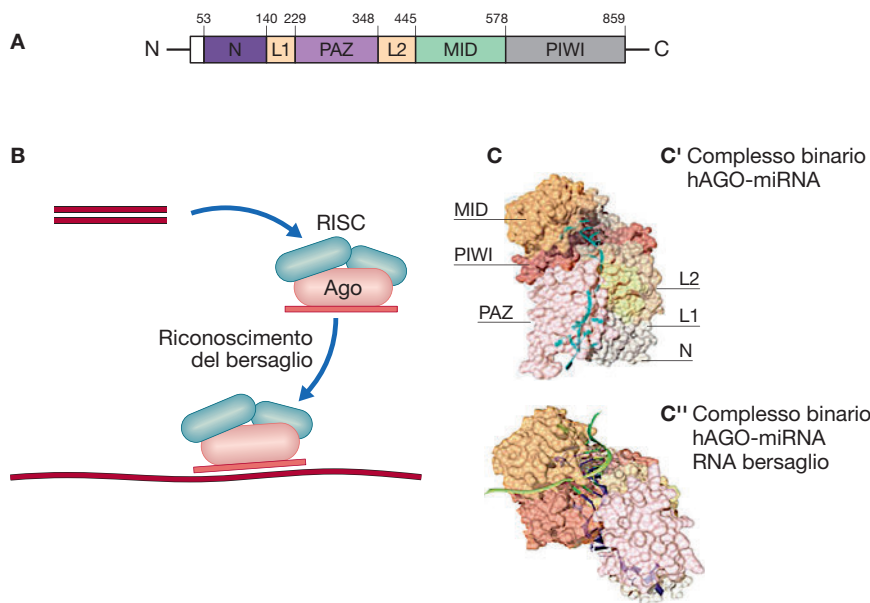


Figura 18.5 La proteina Ago e il complesso RISC. (A) Rappresentazione lineare delle proteine Ago con i suoi domini funzionali N, PAZ, MID, PIWI e i linker L1 e L2. (B) Assemblaggio del complesso RISC. Il miRNA maturo a doppia elica si unisce a RISC con la proteina Ago e viene eliminato il filamento "passeggero". (C) I complessi binario e ternario: nel complesso binario (C') è mostrato l'RNA "guida" (azzurro) che viene presentato con le basi disposte all'esterno per riconoscere l'mRNA bersaglio. Nel complesso ternario (C'') è mostrato l'RNA guida (viola) e l'mRNA target (verde).

RISC è un sofisticato meccanismo di silenziamento genetico e le proteine appartenenti alla famiglia Argonauta (Ago) sono una componente fondamentale e conservata del complesso RISC. L'utilizzo di pochi tipi di proteine Ago permette al complesso RISC di silenziare moltissimi bersagli, senza che la cellula debba sintetizzare a questo scopo numerosissimi complessi enzimatici diversi, questo perché il riconoscimento specifico con il bersaglio è a carico dell'RNA. Come mostrato nella **figura 18.5A**, le proteine Ago contengono quattro domini: N, PAZ, MID, PIWI. AGO assume una forma bilobata costituita dai domini N e PAZ all'estremità N-terminale, e dai domini MID e PIWI all'estremità C-terminale, connessi da due domini di collegamento L1 e L2. Il dominio PIWI assume una conformazione simile alla ribonucleasi H e contiene il sito catalitico coinvolto nella degradazione del bersaglio mediata dai miRNA. Il dominio N, invece, è molto importante per la separazione dei due filamenti del miRNA duplex in seguito al suo caricamento. Mediante studi strutturali, tra i due lobi della proteina è stato individuato un canale in grado di legare acidi nucleici in modo che l'estremità 5' del filamento guida (quello che si appaierà con il bersaglio) sia ancorato al dominio MID di Ago, mentre l'estremità 3' al dominio PAZ. Invece, nella regione 13-16 nt, il filamento guida è ancorato al canale tra i domini N e PAZ, con le basi direzionate verso l'interno e non disponibili per l'appaiamento (**fig. 18.5B e C**). Dopo la formazione del complesso ternario con l'mRNA target, tali nucleotidi cambiano conformazione ed espongono le loro basi verso l'esterno. Questi cambiamenti conformazionali hanno suggerito un possibile modello a più fasi per il riconoscimento di un bersaglio da parte dei miRNA. Dapprima, il miRNA interroga il messaggero mediante appaiamento con la *seed region* (2-5 nt). La sequenza *seed* (seme) è essenziale per il legame del miRNA all'mRNA. Questa regione conservata è composta da sette basi, situate in posizione 2-7 dall'estremità 5' del miRNA. Anche se l'appaiamento tra le basi del miRNA e dell'mRNA target non è perfetto, la *seed sequence* deve essere perfettamente complementare. Questo provoca cambiamenti conformazionali che coinvolgono un' α -elica e il dominio PAZ della proteina Ago, portando all'esposizione dei nucleotidi 6-8 della *seed region* del miRNA e il segmento coi nucleotidi 13-16, per un ulteriore riconoscimento. Inoltre, interazioni con il solco minore del du-

plex miRNA-mRNA permette ad Ago2 di interrogare l'RNA target in modo sequenza indipendente.

L'appaiamento di basi tra il miRNA e l'mRNA bersaglio può essere più o meno perfetto con diversi risultati sul meccanismo di regolazione. Infatti, sembra che si possa generalizzare la conclusione che i miRNA che si appaiano in maniera imperfetta nella regione 3'UTR dell'mRNA bersaglio risultano in un blocco della traduzione dell'mRNA stesso (**fig. 18.4**); al contrario i miRNA che si appaiano perfettamente a sequenze bersaglio che si trovino in qualsiasi posizione lungo l'mRNA, perlopiù nella regione codificante, lo avviano alla degradazione (**fig. 18.4**). Come abbiamo già detto, in genere più miRNA sono coinvolti nella regolazione di un mRNA. C'è ancora molta incertezza su molti aspetti dei meccanismi di queste regolazioni; ad esempio, non è ancora chiaro se la repressione traduzionale sia dovuta a un blocco dell'allungamento della traduzione o dell'inizio, oppure se diversi meccanismi siano implicati nelle varie situazioni. Comunque, si tratta di un controllo negativo, che però è irreversibile nel caso di appaiamenti perfetti che portano a degradazione dell'mRNA e possibilmente reversibile nel caso di interazioni miRNA/mRNA non perfette che reprimono la traduzione lasciando tuttavia la possibilità che questa venga ripresa.

Uno dei bersagli del complesso miRNA-RISC è il fattore eucariotico di inizio della traduzione eIF4A. Questa proteina è una RNA elicasi che svolge le strutture secondarie all'interno della regione 5'UTR (*untranslated region*) dell'mRNA. Ha quindi lo scopo di permettere lo scorrimento della subunità 40S del ribosoma associata a fattori d'inizio della traduzione, di scorrere la regione 5'UTR in cerca del codone d'inizio AUG (*vedi par. 15.2*). Pertanto, l'interazione di RISC con il fattore d'inizio eIF4A, riduce o inibisce la sintesi della proteina codificata dall'mRNA bersaglio, interferendo con la reazione di scanning richiesta per il riconoscimento del codone AUG dell'mRNA.

Questi studi sulla regolazione genica da parte dei miRNA hanno dato il via allo sviluppo di nuove tecniche che hanno grande interesse sia per la ricerca di base che per le loro applicazioni di tipo terapeutico. Infatti, sono stati sviluppati vari metodi di silenziamento genico che vanno sotto il nome di "interferenza da RNA" (*RNA interference*, RNAi), illustrati nella parte destra della **figura 18.4** e di cui parleremo però nel paragrafo 20.10.

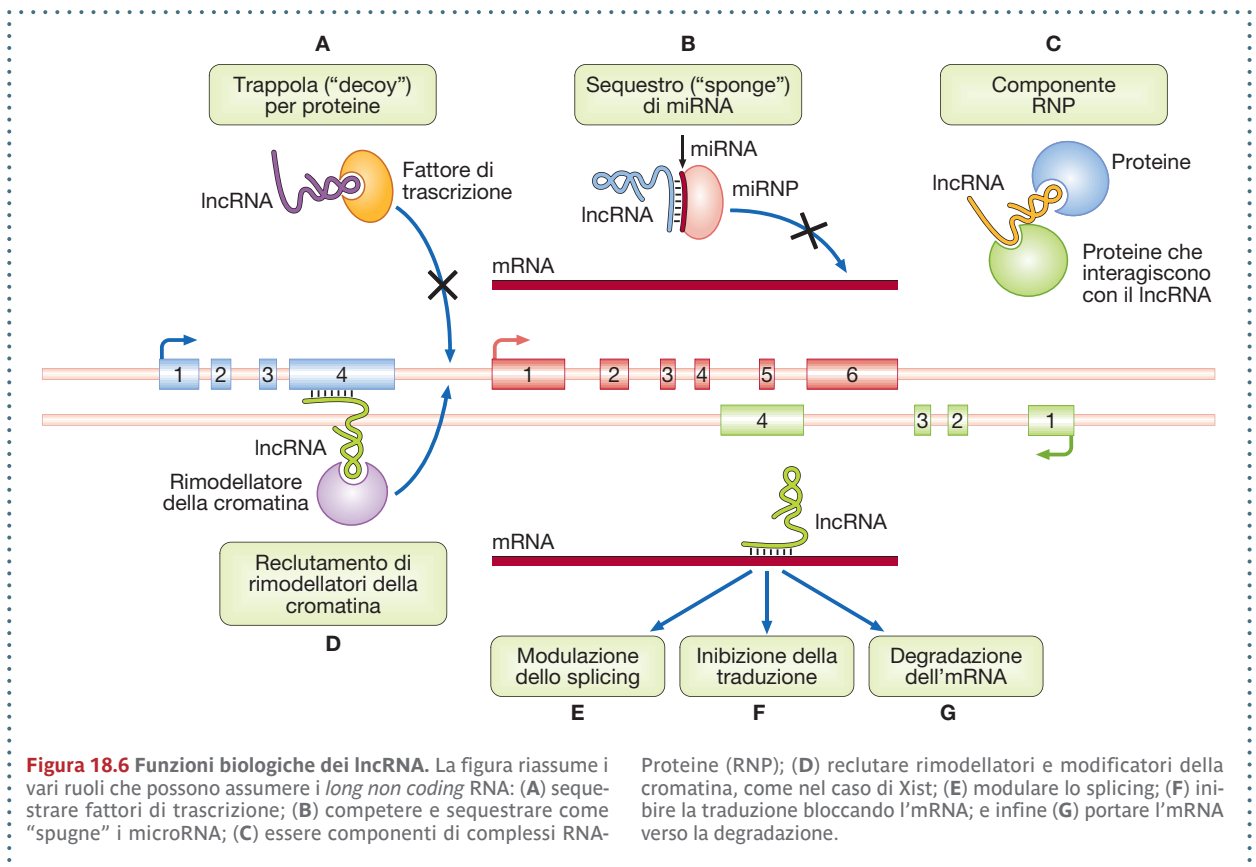
Lunghi RNA non codificanti (lncRNA)

Si è potuto stabilire che una porzione molto ampia dei trascritti delle cellule di diversi organismi eucariotici è costituita da molecole più lunghe di 200 nt che spesso hanno il CAP e sono poliadenilate, ma non contengono evidenti *Open Reading Frames* (ORF) codificanti per proteine: tali molecole di RNA sono definite **long non coding RNA (lncRNA)**. La maggior parte dei trascritti negli eucarioti sono lncRNA, espressi tuttavia singolarmente a livelli molto inferiori rispetto ai trascritti canonici di geni codificanti proteine. Una plausibile interpretazione suggerisce che le RNA polimerasi e i meccanismi di selezione dei promotori agiscano come sistemi non particolarmente stringenti. Anche se le polimerasi trascrivono preferenzialmente un solo filamento del DNA e iniziano, di solito, a specifici TSS (*Transcription Start Sites*), il processo trascrizionale può iniziare, più o meno raramente, anche a una certa distanza dal promotore, eventualmente in direzione opposta, oppure su siti che, pur non essendo normali promotori, presentino una qualche somiglianza con essi e, quindi, una certa propensione a iniziare la trascrizione. Molti lncRNA sono espressi da pseudogeni e molti possono essere il risultato di mutazioni in geni strutturali, duplicazioni, inversioni e inserzioni nel genoma. Portando questa visione all'estremo, si potrebbe pensare che i lncRNA rappresentino un "rumore di fondo" e che non abbiano specifici ruoli funzionali. È stato, invece,

dimostrato che molti lncRNA svolgono importanti ruoli regolativi, cosa d'altra parte plausibile se ipotizziamo che, nei tempi evolutivi, la selezione può aver trovato vantaggioso utilizzare questi RNA per ruoli specifici nella regolazione fine dell'espressione del genoma.

I lncRNA possono svolgere la loro funzione regolativa legando proteine coinvolte in meccanismi di regolazione dell'espressione genica o interagendo direttamente con specifici bersagli rappresentati da molecole di DNA e/o RNA. Inoltre, l'analisi quantitativa dei lncRNA ha stabilito che i loro livelli possono variare in tessuti diversi o durante il differenziamento cellulare, suggerendo che i lncRNA possano svolgere funzioni molto importanti nei meccanismi molecolari di differenziamento e sviluppo.

I lncRNA possono agire sul controllo dell'espressione genica attraverso meccanismi diversi e tale diversità è principalmente correlata alla possibilità che i lncRNA svolgano la loro funzione nel nucleo o nel citoplasma. Nel nucleo la funzione regolativa dei lncRNA si attua principalmente mediante vari meccanismi, alcuni dei quali sono illustrati nella **figura 18.6**. I lncRNA possono fungere da esche (A nella figura) che legano e sequestrano proteine regolative e fattori di trascrizione, impedendo loro di interagire con il DNA. Alternativamente, possono svolgere il ruolo di adattatori o "scaffold" (C nella figura) che mediano la formazione di complessi ribonucleoproteici (RNP) che possono agire anche come modificatori della cromatina, condensandola o aprendola, con conseguente inibizione o attivazione della trascrizione. Esempi rappresentativi di questi complessi di ribonucleoproteine (RNP) sono il Polycomb Repressive Complex (PRC) che silenzia i geni bersaglio portando la cromatina in uno stato condensato e il complesso MLL1 che, al contrario, attiva la trascrizione dei geni bersaglio. Molti lncRNA agiscono da "guida" (D nella figura), guidando la localizzazione su siti bersaglio



nel DNA di enzimi modificatori della cromatina, come la DNA metiltransferasi 3 (DNMT3) che, inducendo la metilazione del DNA ne inibisce la trascrizione genica. I lncRNA possono anche legarsi direttamente all'RNA bersaglio, interferendo così con il meccanismo di splicing, inibendone la traduzione o avviandolo alla degradazione (rispettivamente E, F e G nella figura); in alcuni casi, la semplice sintesi di un lncRNA può interferire negativamente con l'espressione genica, ad esempio ostacolando l'azione dell'RNA polimerasi.

Considerando la versatilità funzionale dei lncRNA, un singolo trascritto potrà fungere allo stesso tempo sia da scaffold che da guida. In base al sito bersaglio, i lncRNA guida possono anche essere suddivisi in due categorie: lncRNA che agiscono in *cis* o in *trans*. I primi controllano l'espressione di geni posizionati nelle vicinanze del sito in cui il lncRNA viene trascritto, mentre i secondi attivano o reprimono l'espressione di geni in loci indipendenti. Attraverso modificazioni della cromatina alcuni lncRNA sono anche in grado di attivare specifiche sequenze enhancer che abbiamo già ampiamente descritto nel paragrafo 11.5.

Come esempio importante del ruolo di un lncRNA che agisce nel nucleo modulando la conformazione della cromatina possiamo brevemente descrivere la funzione di Xist (*X-inactive specific transcript*), un lncRNA responsabile dell'inattivazione di un intero cromosoma X. Xist è espresso esclusivamente nelle cellule dove sono presenti due cromosomi X ed è trascritto nella regione Xic (*X chromosome inactivation center*) del cromosoma che deve essere inattivato. Questo lncRNA è lungo circa 17 kb e si lega al cromosoma da cui è trascritto richiamando fattori di rimodellamento della cromatina che silenziano completamente il cromosoma. Il suo antagonista che impedisce all'altro cromosoma X di essere inattivato è Tsix, che è anch'esso un lncRNA, trascritto in direzione antisenso a Xist. Tsix, come Xist, agisce esclusivamente sul cromosoma che lo produce legandosi in maniera complementare a Xist, inibendo così la sua azione di silenziamento. Un'aberrante espressione di Xist è stata osservata in alcuni tumori dell'uomo, indicando come alterazioni nel processo di inattivazione di un cromosoma X possano contribuire all'insorgenza di tumori.

Come anticipato sopra, numerosi lncRNA esplicano la loro funzione regolatrice nel citoplasma interagendo con sequenze complementari trascritte sullo stesso locus cromosomico o su loci indipendenti. In questo compartimento cellulare un ruolo molto importante di alcuni lncRNA, confermato da numerosi dati sperimentali, è di agire come "spugne" che, sequestrando specifici miRNA, interferiscono con la loro funzione (**fig. 18.6B**). Il meccanismo si basa sul fatto che il lncRNA può contenere una regione complementare al miRNA, analoga al MRE (*miRNA response element*) dell'mRNA bersaglio, capace quindi di "competere" con il miRNA per l'interazione con un mRNA. Questa interazione descrive un nuovo tipo di circuito regolatore dove diversi tipi di RNA possono confrontarsi tra loro e competere per i miRNA che hanno in comune.

Questi lncRNA con funzione di "spugna" che sequestrano specifici miRNA sono anche chiamati ceRNA (*Competitive Endogenous RNA*) (**fig. 18.7A**), e giocano una funzione importante nella regolazione del differenziamento e dello sviluppo così come nel modulare la formazione o soppressione di tumori. Questi lncRNA possono essere pseudogeni, altri messenger che hanno in comune dei miRNA o trascritti senza ORF (**fig. 18.7B**). Recentemente, nelle cellule neuronali, è stata identificata una nuova classe di RNA circolari (circRNA), che, come i ceRNA, possono agire da "spugne" per miRNA ma, al contrario di essi, che avendo un'emivita breve permettono un controllo rapido e variabile della traduzione, i circRNA mostrano una maggiore stabilità e il loro turnover è regolato dall'appaiamento perfetto con un miRNA che causa il taglio endonucleolitico della molecola.

Tutte queste recenti osservazioni sul ruolo dei lncRNA, e più in genere dei ncRNA, nella regolazione fine dell'espressione del genoma suggeriscono che le

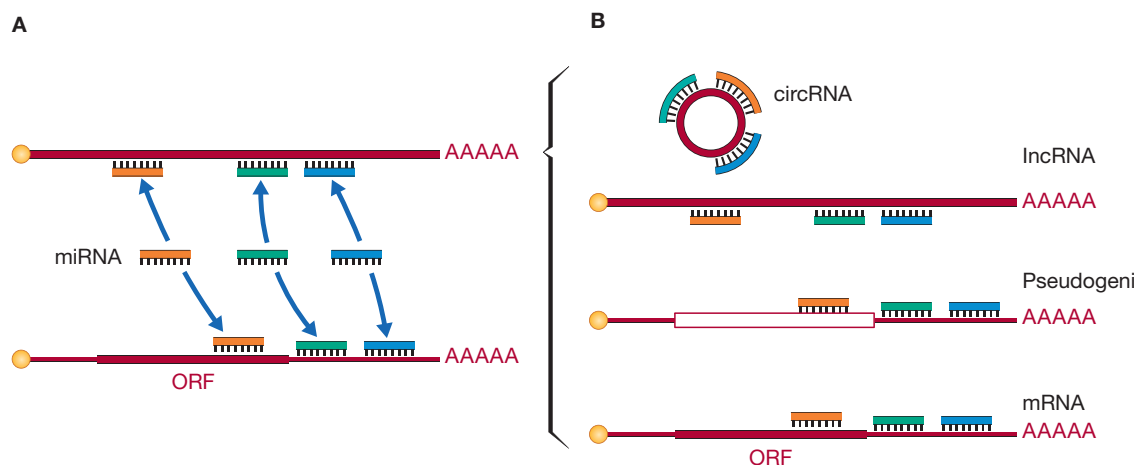


Figura 18.7 Funzionamento dei ceRNA. L'appaiamento tra le basi è alla base del funzionamento dei ceRNA (*competing endogenous RNAs*). (A) mRNA e ncRNA possono legare gli stessi miRNA, e competere per essi. (B) In questi casi la complementarità è tra i miRNA e differenti bersagli tra cui RNA circolari (circRNA), lncRNA, pseudogeni e mRNA. Queste interazioni tra gli RNA che codificano per proteine e quelli non codificanti fa sì che tutti questi possano competere per le sequenze di riconoscimento dei miRNA. (Modificata da: Fatica A. e Bozzoni I. (2014), *Nat. Rev. Genet.* 15, pp. 7-21).

cRNA, pseudogeni e mRNA. Queste interazioni tra gli RNA che codificano per proteine e quelli non codificanti fa sì che tutti questi possano competere per le sequenze di riconoscimento dei miRNA. (Modificata da: Fatica A. e Bozzoni I. (2014), *Nat. Rev. Genet.* 15, pp. 7-21).

molteplici e complicate interazioni tra RNA codificanti e non codificanti generano grandi e complesse reti (network) regolative del trascrittoma. Tutto ciò ci porta a una nuova visione dei meccanismi operanti nella cellula in cui l'interazione tra diversi RNA risulta in una complessa regolazione dell'attività genica.

18.3 Il mondo a RNA e l'origine della vita sulla Terra

La scoperta degli introni capaci di autosplicing in *Tetrahymena thermophila*, fatta da Thomas Cech nel 1982, e degli altri processi sopra descritti, ha rivelato che l'RNA può essere dotato di attività catalitica indipendente dalla presenza di proteine. Questa importante scoperta ha radicalmente cambiato la nostra visione del processo che ha portato all'origine di forme di vita primordiali, avvalorando uno scenario privo di DNA e proteine. Infatti, il progressivo arricchimento delle nostre conoscenze sui processi di base della Biologia Molecolare ha chiaramente svelato il ruolo chiave dell'RNA nelle prime fasi dello sviluppo della vita sulla Terra, che probabilmente videro la formazione di semplici molecole organiche in un'atmosfera non ossidante che, concentrandosi in siti specifici per formare il cosiddetto "brodo primordiale", diedero origine, a fronte di assorbimento di energia, a polimeri organici più complessi, tra i quali l'RNA. L'RNA è dotato sia della capacità di propagarsi attraverso la sua replicazione che di mediare una serie di attività funzionali attraverso le sue proprietà catalitiche. Nelle prime fasi della vita esso avrebbe quindi avuto sia il compito di immagazzinare l'informazione genetica che la capacità di catalizzare reazioni chimiche e avere una funzione strutturale.

È stata così formulata da Walter Gilbert, nel 1986, l'**ipotesi del mondo a RNA**, ossia la vita senza DNA e proteine, che successivamente avrebbero assunto, rispettivamente, il ruolo di depositario dell'informazione genetica e di effettori della sua espressione.

Vi sono numerose osservazioni a supporto di un ruolo preminente di molecole di RNA nei processi vitali, ad esempio:

1. L'attività catalitica necessaria per la sintesi proteica risiede nella componente a RNA del ribosoma. Il ribosoma può essere quindi considerato un ribozima e non si può non considerare realistica la possibilità, come suggerito da Crick nel 1968, che il ribosoma primitivo fosse costituito esclusivamente da RNA.

2. L'RNA messaggero è una molecola chiave nella trasmissione dell'informazione genetica. Molti virus hanno un sistema genetico basato sull'RNA.
3. Il sistema per l'export delle proteine del reticolo endoplasmatico coinvolge un complesso denominato SRP (*Signal Recognition Particle*), di cui una componente essenziale è l'RNA non-codificante 7SL.
4. La telomerasi, richiesta per garantire la fedeltà e completezza del processo replicativo dei cromosomi lineari, utilizza uno stampo a RNA.
5. L'RNAsi P, il complesso proteico coinvolto nella maturazione dell'estremità 5' dei tRNA, ha una componente a RNA che nei batteri è certamente responsabile dell'attività enzimatica.
6. Gli snRNA, i piccoli RNA componenti dello spliceosoma, sono ubiquitari negli eucarioti, anche quelli più primitivi come i microsporidi, collocando l'origine degli introni nelle prime fasi dello sviluppo della vita sulla Terra, come anche testimoniato dagli introni autocatalitici.
7. È probabile che i tRNA pre-datino la sintesi proteica. L'ansa dell'anticodone sembra un'acquisizione più recente, dato che alcuni tRNA hanno uno specifico ruolo catalitico, come la formazione del delta-amminolevulinato nella biosintesi dell'eme nelle piante.
8. Molti coenzimi contengono componenti nucleotidiche (non deossinucleotidiche), quali NAD, FAD, CoA, S-adenosilmetionina, vitamina B₁₂.
9. L'ATP, la molecola bioenergetica per eccellenza, è un ribonucleotide.

Il fatto che l'RNA abbia preceduto l'origine del DNA e delle proteine non esclude la possibilità che macromolecole differenti siano apparse prima.

Rimane da comprendere come si siano generate, attraverso un processo non enzimatico, le prime catene polinucleotidiche e come queste si siano evolute per acquisire diverse proprietà catalitiche, particolarmente quella di autoreplicarsi in modo stampo-dipendente. Una macromolecola reattiva come l'RNA non è comunque quella ideale per conservare e propagare l'informazione genetica. Per questa ragione il processo evolutivo avrebbe generato il DNA, che, come sappiamo, si differenzia dall'RNA per la presenza del deossiribosio al posto del ribosio, e della timina al posto dell'uracile.

La presenza del 2'OH nel ribosio, assente nel deossiribosio, rende la catena polinucleotidica intrinsecamente instabile per la possibilità di un attacco nucleofilo al 3' fosfato che porta alla rottura del legame fosfodiesterico. Inoltre, la struttura a doppio filamento del DNA ne aumenta in modo significativo la stabilità e lo rende una macromolecola ideale per assolvere al ruolo di depositario dell'informazione genetica.

Anche la presenza della timina al posto dell'uracile conferisce un consistente vantaggio evolutivo. Infatti, la citosina è normalmente soggetta a deaminazione ossidativa spontanea trasformandosi in uracile. Il fatto che il DNA usi timina al posto di uracile consente ai sistemi di riparazione di riconoscere in modo selettivo le basi modificate, garantendo quindi una maggiore fedeltà nella trasmissione dell'informazione genetica.

Le potenzialità catalitiche dell'RNA e la varietà di configurazioni strutturali dell'RNA, costituito da soli 4 componenti elementari, sono quindi molto inferiori a quelli delle proteine, che possono formare combinazioni tra 20 diverse unità elementari, a loro volta caratterizzate da diverse proprietà chimico-fisiche e dimensioni. Pertanto, le proteine possono avere una varietà di proprietà catalitiche e strutturali molto maggiore.

In conclusione, l'ipotesi del mondo a RNA apre nuovi orizzonti per comprendere non solo la storia della vita sulla Terra, ma anche i numerosi meccanismi molecolari ancora sconosciuti. È innegabile che l'RNA stia assumendo un ruolo sempre più centrale nella Biologia Molecolare, come testimoniato dalla crescente importanza della funzione svolta dagli RNA non codificanti, sia a basso peso

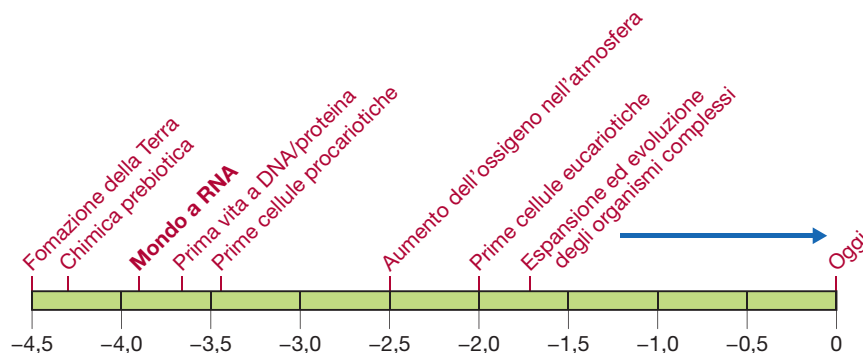
molecolare (miRNA, siRNA) che a elevato peso molecolare, in molti processi biologici fondamentali, inclusa la regolazione del piano di espressione genica in diverse condizioni fisiopatologiche attraverso la modulazione delle proprietà epigenetiche della cromatina (ad es. l'inattivazione di una delle copie del cromosoma X che coinvolge diverse molecole di RNA, tra le quali Xist, un grande RNA non codificante di circa 20 kb).

L'ipotesi del mondo a RNA è plausibile e avvalorata da quanto discusso sopra; tuttavia, non sappiamo con certezza come tutto abbia avuto origine. Per studiare l'evoluzione degli organismi sul nostro pianeta si è fatto fino ad ora uso dei fossili che ci hanno permesso di studiare, in una scala temporale, molti organismi che hanno abitato la Terra. Con le intuizioni fondamentali di C. Darwin abbiamo anche capito come la "selezione naturale del più adatto" sia stato il meccanismo che ha generato tutta la biodiversità esistente nel nostro pianeta partendo da un ipotetico LUCA (*Last Universal Common Ancestor*), l'ultimo e universale nostro antenato comune. Lo studio dei genomi degli organismi animali e vegetali presenti sulla Terra ci ha permesso di disegnare alberi genealogici sempre più precisi e collocare nella giusta casella temporale la comparsa delle varie specie. Tuttavia è impossibile trovare il fossile dell'origine della vita. La definizione di vita descrive l'origine di un processo biologico autoreplicante e che è soggetto all'evoluzione darwiniana.

Tuttavia il famoso esperimento di Miller-Urey eseguito nel 1953 (lo stesso anno della risoluzione della struttura del DNA) ha dimostrato che è possibile disegnare approcci sperimentali *in vitro*, che simulano l'atmosfera terrestre prebiotica e che, in presenza di scariche elettriche, che mimano i fulmini e aumentano l'energia del sistema, è possibile formare molecole organiche, come amminoacidi e precursori delle basi degli acidi nucleici. Su questa linea sperimentale e con le informazioni più dettagliate che abbiamo dell'ambiente prebiotico grazie alle analisi che ci vengono dall'astrobiologia, sappiamo che la polvere cosmica è formata in larga parte dagli elementi che sono alla base dei composti organici, elementi come il carbonio, l'azoto, il fosforo, e naturalmente idrogeno e ossigeno. In particolari condizioni chimico-fisiche, nelle nubi interstellari, il carbonio reagisce con l'azoto e l'idrogeno e si formano in abbondanza l'acido cianidrico HCN e il cianamide H_2N-CN , che sono molecole estremamente reattive. In presenza di acqua dall'acido cianidrico si forma facilmente la formammide NH_2COH che, se scaldata a temperature elevate (170-190 °C) può dare origine a purine da cui, con reazioni chimiche relativamente semplici, possono derivare le quattro basi azotate presenti negli attuali acidi nucleici. Le reazioni di sintesi delle basi azotate a partire da formammide sono molto più efficienti in presenza di argille del tipo montmormilloniti (il nome deriva dalla località di Montmorillon, in Francia), perché la struttura regolare e la capacità assorbente di quel tipo di argille aumenta in maniera significativa la reattività delle molecole agendo come concentratore di sostanze che possono così interagire più facilmente.

Quello che abbiamo brevemente descritto sulla chimica prebiotica può essere avvenuto nello spazio o sulla Terra in quello che Darwin ha definito "*warm little pond*" (piccola pozza calda) e, infatti, molecole organiche sono state trovate su meteoriti e comete. Come sappiamo, le molecole di RNA sono polimeri di nucleotidi e nelle cellule ci sono numerosi enzimi coinvolti nella loro sintesi, ma nel "brodo primordiale" questi enzimi non c'erano. Tuttavia c'erano diversi tipi di minerali argillosi tipo le montmormilloniti citate sopra e, nel 1996, il biochimico Leslie Orgel (uno dei membri del "club della cravatta a RNA"; vedi fig. 2F2.1) ha dimostrato che quando nucleotidi attivati venivano aggiunti a minerali come la montmormillonite, si ottenevano molecole di RNA lunghe fino a 55 nucleotidi. Questi dati sulla chimica prebiotica, la sua relativa sem-

Figura 18.8 La vita sulla Terra. Scala temporale che mostra le varie fasi dell'evoluzione della vita sul nostro pianeta, dalla formazione della Terra ai nostri giorni. Si noti come il mondo a RNA si sia sviluppato in un periodo precoce e relativamente breve.



plicità e l'abbondanza nello spazio dei precursori delle molecole biologiche, ci suggeriscono che il “processo della vita” possa essersi formato con relativa facilità e questo spiega, guardando nella **figura 18.8** la scala temporale dei vari passaggi dall'origine della Terra (presumibilmente 4,5 miliardi di anni fa) fino ai giorni nostri, come sia stata relativamente veloce la formazione del mondo a RNA (meno di 1 miliardo di anni), rispetto agli oltre 3 miliardi di anni per arrivare alla biodiversità attuale.

Lecture di approfondimento

- Fabian, M.R., Sonenberg, N. e Filipowicz, W. (2010), «Regulation of mRNA translation and stability by microRNAs», *Annual Review of Biochemistry*, **79**, pp. 351-379.
- Neeraja, S. (2016), «The RNA World at Thirty: A Look Back with its Author», *J. Mol. Evol.* **83**, pp. 169-175.
- Quinn, J.J e Chang, H.Y. (2016), «Unique features of long non-coding RNA biogenesis and function», *Nat. Rev. Genet.* **17**, pp. 47-62.
- Salmena, L., Poliseno L., Tay Y., Kats L. e Pandolfi, PP. (2011), «A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? », *Cell*, **146**, pp353-358.
- Scott, W.G., Martick, M. e Chi, Y.I. (2009), «Structure and function of regulatory RNA elements: ribozymes that regulate gene expression», *Biochim. Biophys. Acta.* **1789**, pp. 634-641.
- Tezuka, T., Hara, H., Ohnishi, Y., e Horinouchi, S. (2009), «Identification and gene disruption of small noncoding RNAs in *Streptomyces griseus*», *J Bacteriol.* **191**, pp. 4896-4904.

19

Modificazioni e regolazioni post-traduzionali di proteine

- 19.1 Stabilità e processamento di proteine
- 19.2 Ubiquitinazione di proteine
- 19.3 Sumoilazione di proteine
- 19.4 Fosforilazione e defosforilazione di proteine
- 19.5 Acetilazione di proteine
- 19.6 Metilazione di proteine
- 19.7 Glicosilazione e modificazioni lipidiche di proteine

Il DNA è il depositario dell'informazione genetica, ma le proteine sono le maggiori responsabili della struttura e funzione di una cellula. In questo testo abbiamo descritto più volte le funzioni di singole proteine o delle “macchine proteiche” che agiscono in tutti gli aspetti del metabolismo cellulare. Non abbiamo, però, trattato in modo sistematico la composizione e struttura delle proteine, dato che tale problematica è affrontata in un corso di Biochimica ed esistono ottimi testi didattici per tale disciplina.

Riteniamo, però, che sia importante discutere brevemente delle modificazioni post-traduzionali di proteine e di come tali modificazioni costituiscano l'ultimo livello del flusso dell'informazione genetica e della sua regolazione. Esse aumentano la diversità dei gruppi funzionali nelle proteine ben al di là dei 20-22 tipi di amminoacidi esistenti. Tali modificazioni variano le proprietà chimiche di proteine, rappresentano nuovi segnali di riconoscimento molecolare, “accendono” o “spengono” attività enzimatiche, controllano la stabilità e la localizzazione di proteine all'interno della cellula. Le modificazioni post-traduzionali sono tra i meccanismi più importanti per aumentare di uno o due ordini di grandezza il numero di varianti proteiche nella cellula rispetto al numero di geni esistenti nel genoma.

L'insieme di tutte le proteine di una cellula è chiamato **proteoma**. Assumendo che nel genoma umano possano esserci più o meno 25 000 geni trascritti in RNA e tradotti in proteina, si ritiene che il proteoma dell'uomo possa essere 10-100 volte più ampio raggiungendo la complessità di $2,5 \times 10^5$ - $2,5 \times 10^6$ isoforme proteiche diverse. Tale diversità si raggiunge attraverso 3 meccanismi: il primo, ma meno frequente, consiste nell'utilizzo di diverse metionine per l'inizio della traduzione; il secondo riguarda, come abbiamo visto, alcuni meccanismi di processamento post-trascrizionale dell'RNA, quali lo splicing alternativo e l'editing; il terzo riguarda proprio le modificazioni post-traduzionali di proteine. Va anche sottolineato che le modificazioni post-traduzionali con effetto regolativo costi-

tuiscono il meccanismo di regolazione più rapido, dato che tali modificazioni agiscono direttamente sul prodotto genico e non su aspetti più a monte dell'espressione di un gene, quali la trascrizione, il processamento di RNA o la sua traduzione, che richiedono più tempo per essere realizzati.

Una descrizione dettagliata della chimica delle varie modificazioni post-traduzionali e delle loro conseguenze va al di là dello scopo di questo testo e ci limiteremo a descrivere quelle modificazioni che, arbitrariamente, riteniamo più rilevanti nel contesto del controllo dell'espressione genica e dei meccanismi che abbiamo trattato in questo libro.

È necessario ricordare che le modificazioni di proteine possono essere reversibili o irreversibili, fisiologiche o patologiche. Di queste ultime non parleremo, ma sono rilevanti e sono causa di gravi patologie nell'uomo. Per esempio, le proteine sono suscettibili a varie modificazioni ossidative causate da specie reattive dell'ossigeno (ROS) o dell'azoto (RNS), che includono sia radicali primari ($\bullet\text{OH}$, $\text{O}_2\bullet^-$, $\text{NO}\bullet$) che specie non-radicali (H_2O_2 , CO , N_2O_2 , NO_2 ecc.). È noto che il livello di proteine ossidate aumenta con l'età e nello sviluppo di numerose malattie e, in genere, la modificazione ossidativa di una proteina porta a una diminuzione o perdita della sua funzione biologica.

19.1 Stabilità e processamento di proteine

Alcune modificazioni post-traduzionali sono irreversibili e l'esempio più classico consiste nel processamento di proteine tramite rottura del legame peptidico catalizzata da **proteasi**, enzimi che degradano altre proteine. L'attivazione di proenzimi tramite proteolisi specifica è comune nella produzione di enzimi digestivi, di ormoni, di proteine coinvolte nella coagulazione del sangue e in numerosi altri processi.

In **figura 19.1** è mostrata, a titolo esemplificativo, la conversione di un precursore proteico in una molecola attiva dell'**ormone insulina** mediante tagli proteolitici successivi. La preproinsulina contiene all'estremità aminotermine un tratto idrofobico (chiamato **sequenza segnale**) e tre regioni, chiamate A, B e C. La sequenza segnale dirige la proteina nel lume del reticolo endoplasmatico, dove è convertita in proinsulina, mediante eliminazione proteolitica della sequenza segnale. La proinsulina è poi trasportata nell'apparato di Golgi e nei granuli secretori, dove la regione C di connessione è anch'essa eliminata e i due segmenti A e B sono connessi tra loro, tramite due legami disolfuro, a costituire l'insulina matura che viene secreta.

Un'altra modificazione, la poliubiquitinazione di cui discuteremo poco più avanti, dirige invece molte proteine verso un vero e proprio organello cellulare deputato alla degradazione di proteine e chiamato **proteosoma**.

Il proteosoma (**fig. 19.2**) è un complesso proteico di grosse dimensioni (normalmente 26S): è costituito da una parte centrale (*core particle* di 20S) e da due complessi regolativi (*regulatory caps* di 19S) sormontati da un cappuccio o lid. Il core è formato, nei diversi organismi, da un numero variabile di proteine (circa 30), mentre il cap è costituito da 19 polipeptidi. L'assemblaggio del proteosoma è un processo complesso che richiede ATP. Alcune delle proteine che costituiscono il lid legano selettivamente i substrati da distruggere. Proteine del cap che appartengono alla famiglia delle AAA-proteasi, tramite l'idrolisi di ATP, svolgono le proteine da degradare distruggendo la loro struttura terziaria (folding) e le avviano verso la camera di degradazione vera e propria che si trova nella *core particle* del proteosoma (**fig. 19.3**).

Il proteosoma è presente in molte copie per cellula che si trovano sia nel citoplasma che nel nucleo e in organelli, quale il reticolo endoplasmatico. Il proteosoma non degrada soltanto le proteine poliubiquitinate, ma anche quelle che sono

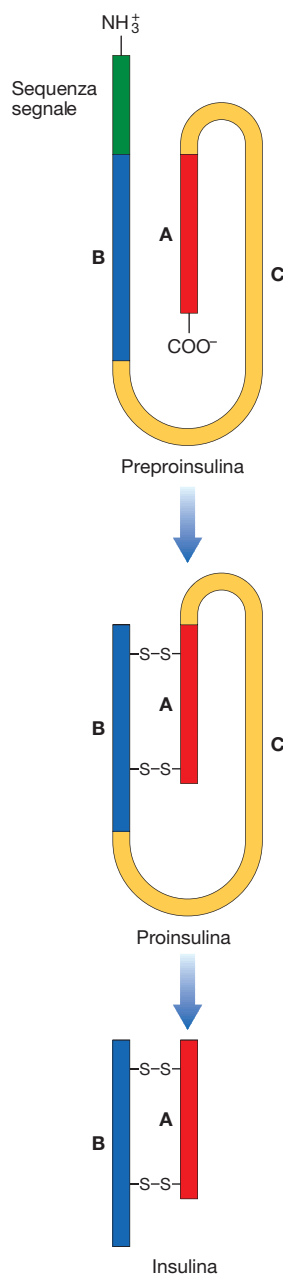


Figura 19.1 Conversione proteolitica del precursore preproinsulina in insulina matura.

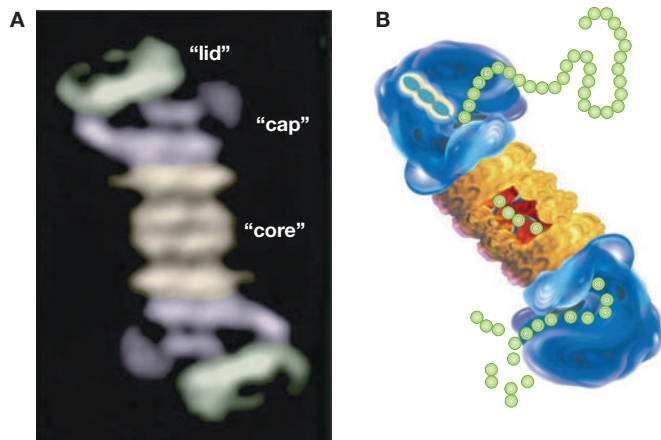


Figura 19.2 Struttura del proteosoma. (A) La parte centrale (in giallo) rappresenta il "core" proteolitico dell'enzima (20S) che è in contatto, sia sopra che sotto, con una "base" o "cap" (19S, in violetto) che contiene una porzione terminale chiamata "lid" o cappuccio. L'immagine è stata generata tramite l'integrazione di diverse fotografie di proteosomi di *Xenopus laevis* ottenute con il microscopio elettronico. (Modificata da: Mcmillan Publishers Ltd: *Journal of Structural Biology*, 1998.) (B) Visione schematica dell'intero proteosoma.

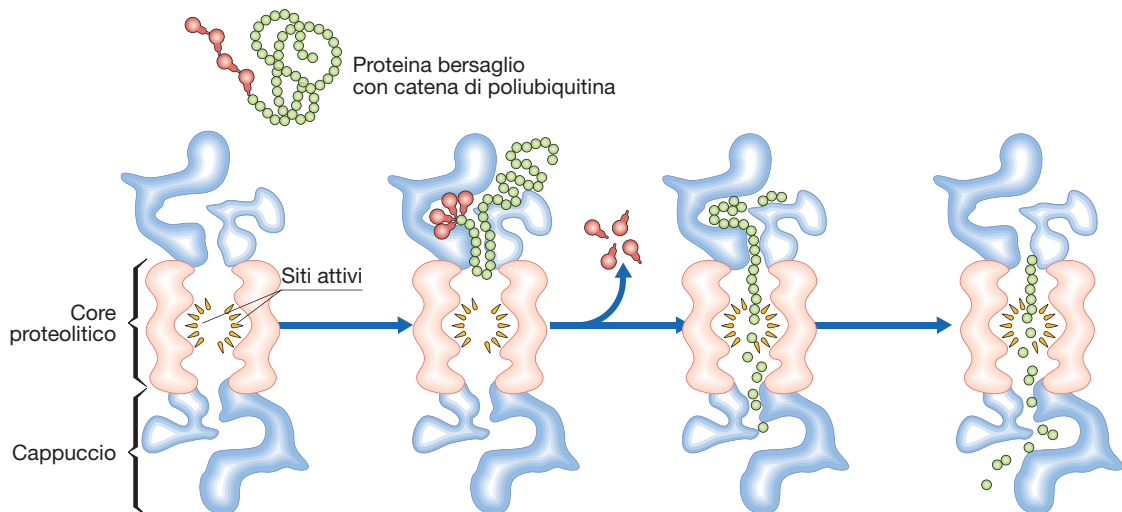


Figura 19.3 Digestione di una proteina da parte del proteosoma. Il "lid" del proteosoma riconosce la proteina da degradare, per lo più marcata con poliubiquitina; successivamente, le AAA-

proteasi del "cap" eliminano il folding della proteina e la spingono nella camera di digestione presente nel core del proteosoma da cui poi fuoriescono peptidi di piccole dimensioni.

riconosciute come ripiegate in una struttura tridimensionale non corretta. Una proteina mal strutturata non solo è inutile per la cellula, in quanto non è in grado di svolgere la funzione biologica richiesta, ma è anche pericolosa. Per esempio, le proteine mature non presentano di solito lunghe regioni di amminoacidi idrofobici sulla superficie esterna: tali regioni possono determinare, infatti, l'aggregazione di proteine non correttamente strutturate che può risultare molto dannosa. Regioni idrofobiche possono formarsi naturalmente durante il processo di ripiegamento (folding) di una proteina, ma sono normalmente mascherate da altre proteine che hanno proprio il compito di facilitare il corretto folding durante la sintesi di una proteina. Tali proteine che guidano il corretto ripiegamento di altre proteine prendono il nome di **chaperoni molecolari**. Esistono numerosi "chaperoni" con capacità di sorveglianza e correzione del folding: in **figura 19.4** è mostrato lo schema di funzionamento dei chaperoni molecolari della famiglia Hsp70. Esiste, quindi, un "controllo di qualità" del folding di una proteina che assomiglia, in qualche modo, al controllo della produzione di un mRNA (*vedi par. 16.3*).

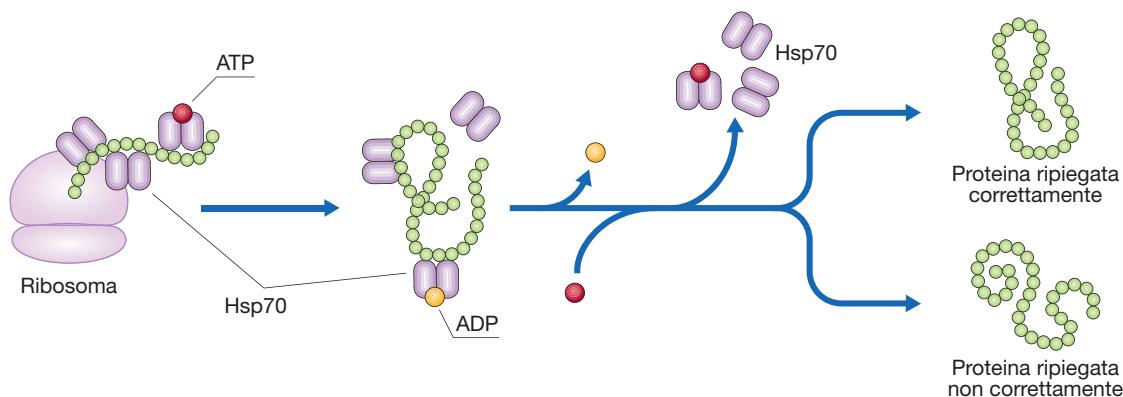


Figura 19.4 Famiglia Hsp70 di chaperoni molecolari. Questi chaperoni molecolari riconoscono le regioni idrofobiche dei loro bersagli già durante la fase di sintesi sul ribosoma. I chaperoni, legandosi alla proteina bersaglio, ne facilitano e controllano il

corretto ripiegamento consumando ATP; i chaperoni sono poi rilasciati al completamento del processo di folding. Le copie di proteine bersaglio non correttamente ripiegate sono avviate alla degradazione tramite il proteosoma.

19.2 Ubiquitinazione di proteine

L'**ubiquitina** è una piccola proteina di 76 amminoacidi molto conservata tra gli eucarioti, con solo 3 amminoacidi diversi tra il lievito e l'uomo (**fig. 19.5**). Come abbiamo detto sopra, il ruolo più noto dell'ubiquitinazione consiste nell'“etichettare” proteine che devono essere indirizzate alla degradazione tramite il proteosoma. Negli ultimi anni si è però scoperto che l'ubiquitinazione di proteine è coinvolta in una varietà di processi, tra cui la proliferazione cellulare, la progressione del ciclo cellulare, l'apoptosi, la riparazione del DNA, il differenziamento, il trasporto di proteine. Tale modificazione è quindi multifunzionale, come la fosforilazione, di cui parleremo più avanti, e con la quale è intimamente interconnessa.

Se l'ubiquitina può essere attaccata a numerose e diverse proteine, sorge spontaneo domandarsi in che cosa consiste la specificità di tale modificazione. Come mai in alcuni casi una proteina ubiquitinata è dirottata verso il proteosoma, mentre in altri casi tale modificazione ha funzioni regolative? Non c'è ancora una risposta definitiva a questo puzzle, ma i pezzi cominciano a incastrarsi cor-

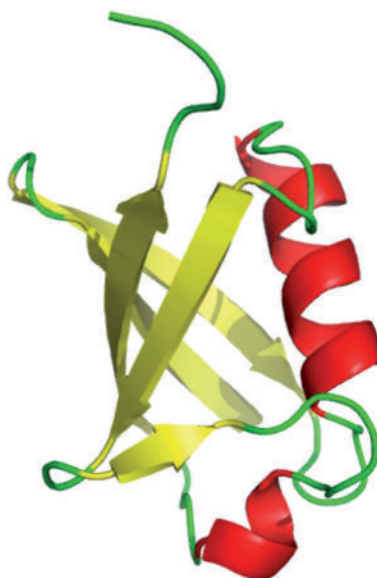


Figura 19.5 Struttura dell'ubiquitina.

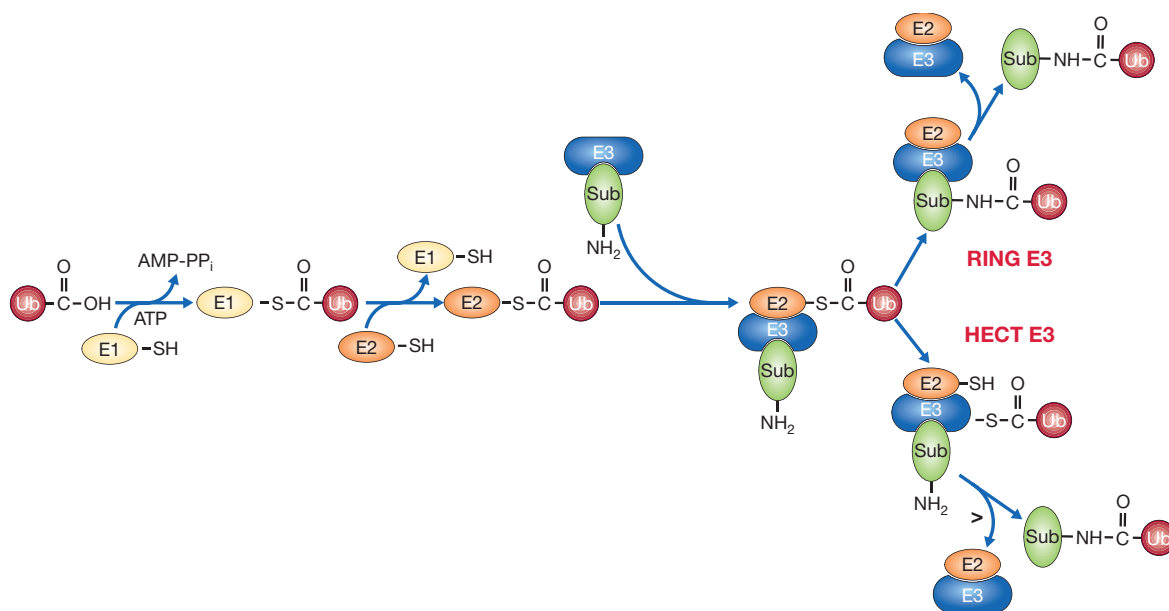


Figura 19.6 Il pathway coinvolto nell'ubiquitinazione di una proteina substrato. L'ubiquitina (Ub) viene attivata utilizzando ATP e formando un estere tioico tra la glicina C-terminale e l'enzima E1. L'ubiquitina può poi essere trasferita a diverse E2, che, a loro volta, possono associarsi con diverse E3. Esistono due famiglie di

proteine E3 chiamate *HECT* e *ring fingers*. Nel caso delle HECT E3, l'ubiquitina viene prima trasferita dall'E2 all'E3 e, successivamente, da questa alla proteina substrato (Sub). Invece, nel caso delle RING E3, l'ubiquitina viene trasferita direttamente dall'E2, ubiquitinata e legato all'E3, alla proteina substrato.

rettamente. Innanzitutto, il legame di una singola molecola di ubiquitina non indirizza la proteina modificata verso la degradazione, mentre la formazione di una catena di 4 o più ubiquitine ha questa funzione. Esistono, inoltre, sottili differenze nel modo in cui una catena di poliubiquitina viene costruita, dato che le diverse molecole di ubiquitina possono legarsi tra loro utilizzando residui diversi di lisina. Infine, il destino delle proteine ubiquitinate può variare a seconda della localizzazione intracellulare in cui avviene tale modificazione. In prima approssimazione, si può dire che la poliubiquitinazione controlla la degradazione di una proteina, mentre la monoubiquitinazione ha funzioni regolative.

L'ubiquitinazione di una proteina è un processo che richiede più passaggi (**fig. 19.6**). Il primo passaggio consiste nell'attivazione dell'ubiquitina stessa, il cui residuo di glicina C-terminale forma un estere con il gruppo SH di una cisteina presente su un enzima, chiamato *ubiquitin activating enzyme* o **E1**. Successivamente, un altro enzima, noto come *ubiquitin conjugating enzyme* (UBC o **E2**), rileva l'ubiquitina da E1 tramite una reazione di transtioilazione che coinvolge ancora la glicina C-terminale. Infine, una terza proteina, nota come *ubiquitin ligase* o **E3**, catalizza il trasferimento dell'ubiquitina dall'E2 all' ϵ -ammino gruppo di una lisina sulla proteina substrato. Esistono due famiglie di proteine E3: la famiglia **HECT** che forma un estere intermedio sulla proteina E3 prima di trasferire l'ubiquitina al bersaglio; la famiglia delle proteine chiamate *ring fingers* che mediano direttamente il trasferimento dell'ubiquitina dall'E2 al substrato. Mentre esiste un solo enzima E1, esistono molti E2 ed E3, ed è proprio la molteplicità di combinazioni tra E2 ed E3 che determina la specificità di substrato della reazione.

Il modo in cui l'ubiquitina viene legata a un substrato ne influenza il suo destino. Una proteina può, infatti, venire modificata su una o più lisine (K) o formare catene di poliubiquitinazione o combinazioni dei due casi (**fig. 19.7**). La scelta della lisina è anche importante nella formazione di catene di poliubiquitinazione: l'ubiquitina contiene, infatti, numerose lisine in posizioni molto

Figura 19.7 L'ubiquitina si lega ad altre proteine e può formare catene di poliubiquitina coinvolgendo residui diversi di lisina.

(A) Le proteine bersaglio possono essere monoubiquitinate su un unico residuo di lisina (K) (AI) o su più residui K diversi (AII). Più ubiquitine possono legarsi tra loro (AIII e AIV) formando varie strutture di poliubiquitinazione a seconda delle lisine coinvolte nell'ubiquitina stessa. (B) Segnali di poliubiquitinazione tramite legami sulla K48 o K29 dell'ubiquitina portano alla degradazione tramite il proteosoma, mentre catene di poliubiquitina formate tramite la K63 svolgono funzioni regolative in altri processi cellulari.

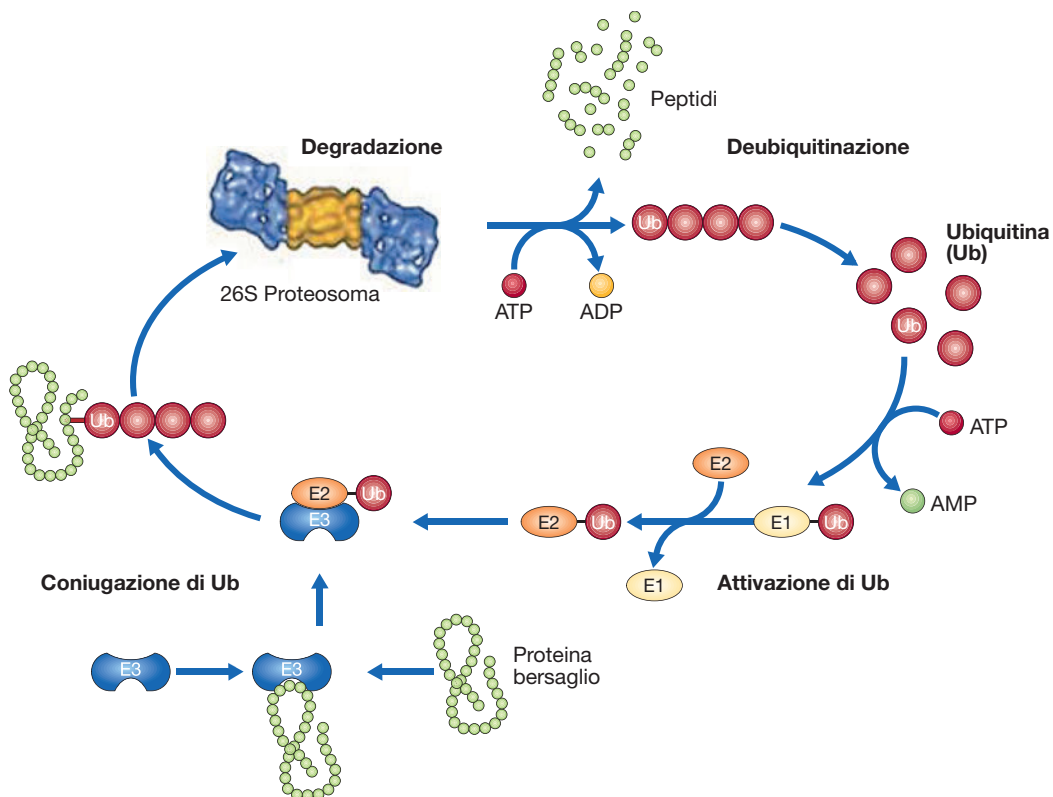
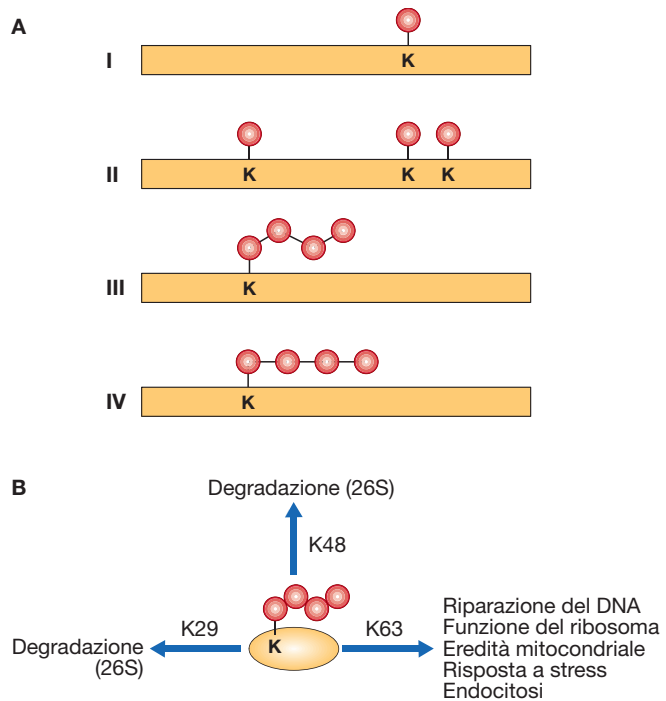


Figura 19.8 Pathway di degradazione di una proteina tramite la sua poliubiquitinazione e proteolisi da parte del proteosoma.

conservate negli eucarioti e tutte sono potenziali siti per la formazione di un legame isopeptidico con il COOH-terminale di un'altra molecola di ubiquitina. Nella cellula, K11, K29, K48 e K63 sono tutti possibili siti per la formazione di legami ubiquitina-ubiquitina e le varie catene di poliubiquitina possono portare diverse informazioni strutturali. È noto che la poliubiquitinazione che si forma tramite legame con la lisina K48 e K29 dirige la proteina poliubiquitinata verso il proteosoma (**fig. 19.8**). I legami, invece, che coinvolgono la lisina K63 hanno, per lo più, funzioni regolative.

Va anche sottolineato che il processo di ubiquitinazione è reversibile e che esistono complessi enzimatici in grado di deubiquitinare proteine; quindi, la regolazione di un certo processo biologico può essere influenzata dall'equilibrio fra attività enzimatiche che controllano in senso positivo o negativo lo stato di ubiquitinazione della proteina bersaglio.

Si potrebbero fare dozzine di esempi importanti su come l'ubiquitinazione di una proteina ne influenza l'attività. Ne faremo uno dopo aver parlato di un'altra modificazione post-traduzionale. Vale la pena anche di ricordare che l'ubiquitinazione è stata recentemente scoperta come una delle modificazioni post-traduzionali degli istoni H2A e H2B e ciò ha importanti funzioni regolative.

19.3 Sumoilazione di proteine

Oltre all'ubiquitina, esistono negli eucarioti altri 10 piccoli peptidi, chiamati **ubiquitin-like proteins** o **Ubl**, che possono legarsi covalentemente ad altre proteine. Tra le Ubl, quella che ha recentemente acquisito sempre più importanza si chiama **SUMO** (*Small Ubiquitin like Modifier*). Nei mammiferi sono stati sino a oggi identificati 4 tipi di SUMO (SUMO1-4), e proteine SUMO esistono in tutti gli organismi eucariotici; le dimensioni delle diverse forme di SUMO sono intorno ai 100 amminoacidi (**fig. 19.9**).

SUMO viene coniugato alle proteine bersaglio seguendo un pathway analogo a quello descritto per il legame dell'ubiquitina e coinvolge enzimi di tipo E1, E2 ed E3 (**fig. 19.10**). Anche in questo caso, la lisina (K) è l'amminoacido che viene sumoilato e la modificazione avviene, preferenzialmente, nell'ambito della sequenza consenso γ KxE (dove γ è un grande residuo idrofobico, K è la lisina, x rappresenta qualsiasi amminoacido ed E è l'acido glutammico).

Studi proteomici hanno individuato sino a oggi circa 130 proteine sumoilate nei mammiferi, un numero certamente destinato ad aumentare. La maggior par-

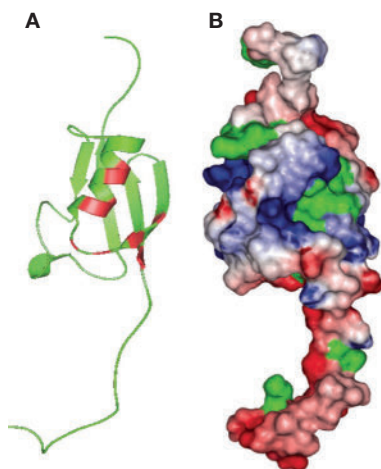
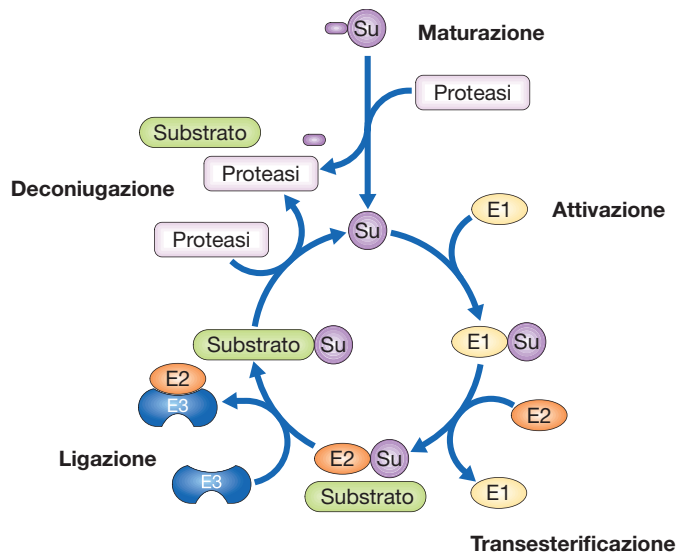


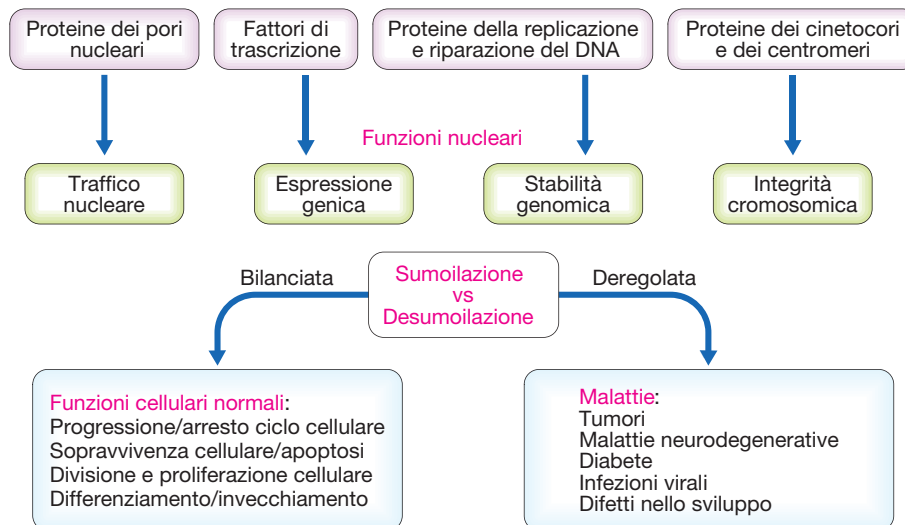
Figura 19.9 Struttura di SUMO. (A) Diagramma a nastro della struttura 3D di SUMO-1. (B) Rappresentazione di superficie della struttura 3D di SUMO-1.

Figura 19.10 Il ciclo di SUMO.

Un precursore di SUMO è processato da una specifica proteasi per esporre al suo C-terminale il residuo di glicina che viene attivato dall'appropriato enzima E1; successivamente SUMO viene trasferito a un enzima E2, chiamato Ubc9. SUMO può essere poi trasferito direttamente sulla proteina bersaglio, o diverse proteine E3 intervengono nel processo facilitando il legame di SUMO al bersaglio. SUMO viene poi staccato dalla proteina bersaglio tramite l'azione di una proteasi.



Substrati di SUMO

**Figura 19.11** Alcune delle varie funzioni di substrati di SUMO.

te di queste proteine è nucleare sottolineando, da una parte, che la sumoilazione gioca un ruolo rilevante in tale comparto cellulare e, dall'altra, che verosimilmente, molte proteine a diversa localizzazione non sono ancora state scoperte, dato che il loro numero è in aumento costante. Come mostrato in **figura 19.11**, un corretto rapporto tra sumoilazione e desumoilazione è estremamente importante poiché un suo sbilanciamento è legato, nell'uomo, all'insorgenza di gravi malattie. La sumoilazione, insieme all'ubiquitinazione, svolge un ruolo molto rilevante in alcuni processi essenziali, quali la riparazione e la replicazione del DNA.

Poiché ubiquitinazione e sumoilazione sono spesso collegate tra loro, riportiamo qui un esempio di come tali modificazioni post-traduzionali regolano la funzione di un'importante proteina, il *Proliferating Cell Nuclear Antigen* o PCNA, che abbiamo incontrato più volte trattando della replicazione e ripara-

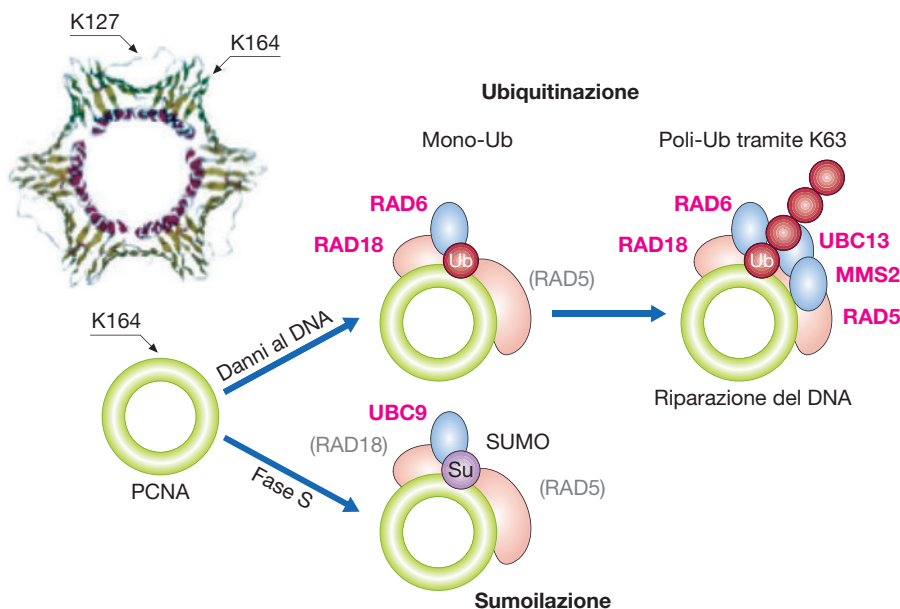


Figura 19.12 Correlazioni tra ubiquitinazione e sumoilazione nella funzione di PCNA. In alto a sinistra, un modello della struttura del PCNA di lievito, dove sono indicate le posizioni delle lisine K127 e K164. Danni al DNA inducono la monoubiquitinazione di PCNA sulla lisina K164 catalizzata dal complesso E2-E3 Rad6 e Rad18. Danni al DNA inducono anche il trasporto nucleare del complesso E2-E3 Ubc13 e Mms2 che viene reclutato sul DNA tra-

mite la proteina Rad5. Tale complesso catalizza la poliubiquitinazione di PCNA tramite la lisina K63. In assenza di danni sul DNA, la proteina PCNA viene sumoilata durante la fase S attraverso l'azione di diverse proteine, tra cui Ubc9. Ubc9 interagisce con il PCNA e, probabilmente, anche con Rad18 e Rad5. I geni codificanti per le proteine che ubiquitinano o sumoilano PCNA sono scritti in rosso, gli altri in grigio o in parentesi.

zione del DNA nei capitoli 6 e 7. PCNA è, infatti, una proteina che interagisce con numerose DNA polimerasi eucariotiche e ne regola la funzione. PCNA è il bersaglio sia di eventi di ubiquitinazione che di sumoilazione, ma l'aspetto più rilevante è che il residuo K164 di PCNA può essere sia monoubiquitinato che monosumoilato, mentre il residuo K127 può venire soltanto sumoilato e tale modificazione, sia su K164 che su K127, avviene durante la fase S. La sumoilazione ha un ruolo replicativo mentre inibisce la riparazione del DNA. Tuttavia, in risposta a danni al DNA, PCNA viene monoubiquitinato su K164 competendo, quindi, con la sumoilazione dello stesso residuo. Alla monoubiquitinazione su K164 da parte del complesso E2-E3 Rad6-Rad18 segue una poliubiquitinazione da parte del complesso E2-E3 Ubc13 e Mms2 tramite legami ubiquitina-ubiquitina che coinvolgono il residuo K63 (**fig. 19.12**). Tale poliubiquitinazione non promuove, però, la degradazione di PCNA, ma, anzi, favorisce il ruolo di PCNA nei meccanismi di riparazione error-free (*vedi* cap. 7).

19.4 Fosforilazione e defosforilazione di proteine

La fosforilazione di proteine è la modificazione post-traduzionale più frequente nel proteoma degli eucarioti. È stato valutato che circa un terzo delle potenziali proteine del proteoma umano è fosforilato e che esistono circa 500 enzimi in grado di fosforilare altre proteine e 150 capaci di defosforilarle. Gli enzimi fosforilanti sono chiamati **proteine chinasi (PK)**, mentre quelli che defosforilano sono denominati **fosfatasi**. La fosforilazione di proteine avviene anche nei procarioti, anche se è estremamente più abbondante negli eucarioti.

Le proteine chinasi vengono classificate in numerose famiglie distinte a seconda della loro struttura e funzione e sarebbe qui impossibile darne una descrizione dettagliata. In estrema sintesi, le proteine chinasi catalizzano il trasferimento del

γ -fosfato dell'ATP su un residuo di serina, treonina o tirosina con liberazione di ADP. A seconda del residuo fosforilato possiamo distinguere le chinasi in 3 categorie: (1) **serina/treonina chinasi**; (2) **tirosina chinasi**; (3) **chinasi a doppia specificità** in grado di fosforilare tutti e tre i residui. La prima categoria è la più abbondante e riflette anche l'abbondanza relativa degli amminoacidi che possono essere fosforilati nel proteoma analizzata mediante spettrometria di massa: l'abbondanza relativa di pS:pT:pY fosforilati misurata in tali studi proteomici è di circa 1800:200:1. Anche se meno frequente, la fosforilazione su tirosina non è meno importante: vedremo, infatti, che svolge funzioni relevantissime nella cellula.

Il sito di legame per l'ATP è generalmente ben conservato in tutte le chinasi, mentre varia notevolmente il sito di riconoscimento tra le chinasi e i loro substrati. L'introduzione di un gruppo fosfato nella struttura terziaria di una proteina crea l'opportunità di creare nuovi appaiamenti elettrostatici e legami idrogeno che possono determinare una riorganizzazione locale della proteina anche molto drammatica. Inoltre, l'introduzione di un gruppo fosfato può essere riconosciuta da altre proteine che hanno dei domini specifici per riconoscere proteine fosforilate. Ne deriva che la fosforilazione è una delle modificazioni post-traduzionali più rilevanti per modulare le interazioni proteina-proteina nella cellula: questo spiega perché la fosforilazione gioca un ruolo così rilevante in praticamente tutte le vie di trasduzione di segnali inter- e intracellulari.

Ci sono 91 **proteine tirosina chinasi (PTK)** nel genoma umano che rappresentano, quindi, circa il 10% del totale delle proteine chinasi e 59 di esse sono dei recettori transmembrana. Le PTK sono classificate in diverse famiglie (**fig. 19.13**) a seconda delle loro somiglianze strutturali; mutazioni nei geni codificanti le PTK sono state identificate come responsabili dell'insorgenza di vari tipi di tumore nell'uomo. La regolazione fisiologica di questi recettori è importante per capire come il loro malfunzionamento possa essere collegato alla cancerogenesi. Infatti, molte PTK innescano vie di trasduzione del segnale all'interno della cellula. Per ottenere ciò, il legame del ligando al recettore ne determina l'oligomerizzazione che è richiesta per l'autofosforilazione delle subunità del recettore su residui di tirosina. Questa modificazione amplifica l'attività chinasi del recettore e media il suo legame con proteine citoplasmatiche in grado di trasdurre il segnale più a valle su altri componenti della cascata di trasduzione del segnale. La sovrapproduzione di tali recettori o riarrangiamenti cromosomici che causano la formazione di proteine di fusione tra una parte del recettore e altre proteine in grado d'indurre una funzione di oligomerizzazione sono stati trovati in molti tumori nell'uomo. Si ritiene che l'attività che causa l'insorgere del fenotipo tumorale sia riconducibile a un'aumentata o costitutiva attività chinasi che porta sia ad alterazioni quantitative che qualitative nella trasduzione del segnale a substrati a valle nella cascata. Le 32 tirosina chinasi che non agiscono come recettori sono localizzate nel citoplasma e il mancato funzionamento di circa la metà di esse è collegato nuovamente all'insorgenza di tumori.

Per quanto riguarda le **serina/treonina proteine chinasi**, anch'esse sono classificate in famiglie distinte, alcune delle quali richiedono per la loro funzione cofattori, come cAMP, ioni calcio ecc. Una descrizione dettagliata di tutte queste famiglie va al di là dello scopo di questo testo, essendo coinvolte essenzialmente in tutti i processi cellulari. La determinazione della struttura ai raggi X del sito catalitico di oltre 30 pS/pT chinasi indica che hanno un'architettura molto conservata. Le chinasi si trovano per lo più nello stato "spento" in assenza di particolari stimoli. Uno dei meccanismi più utilizzati consiste nell'avere dei domini regolativi che possono trovarsi sia in *cis* sulla stessa molecola che in *trans* su una subunità regolativa distinta. Sono stati trovati oltre cento domini che modulano

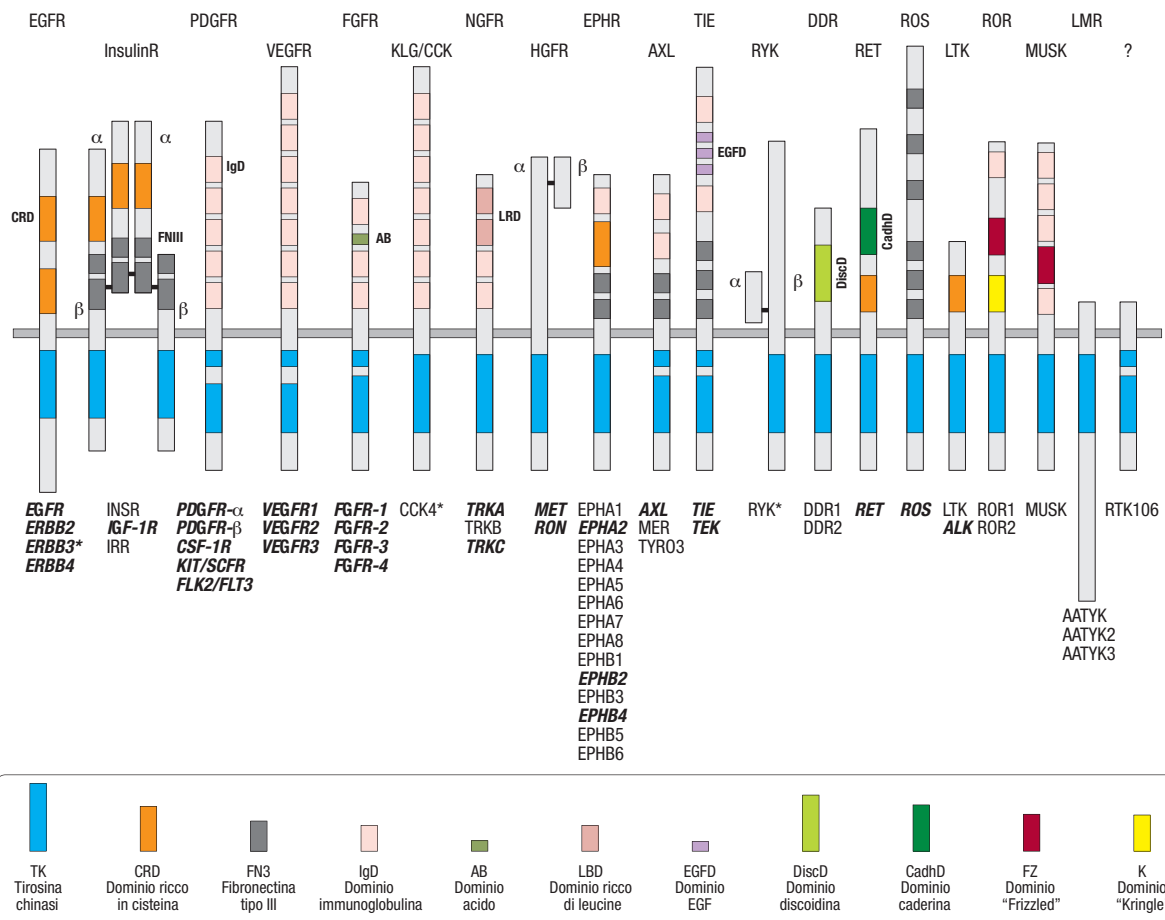
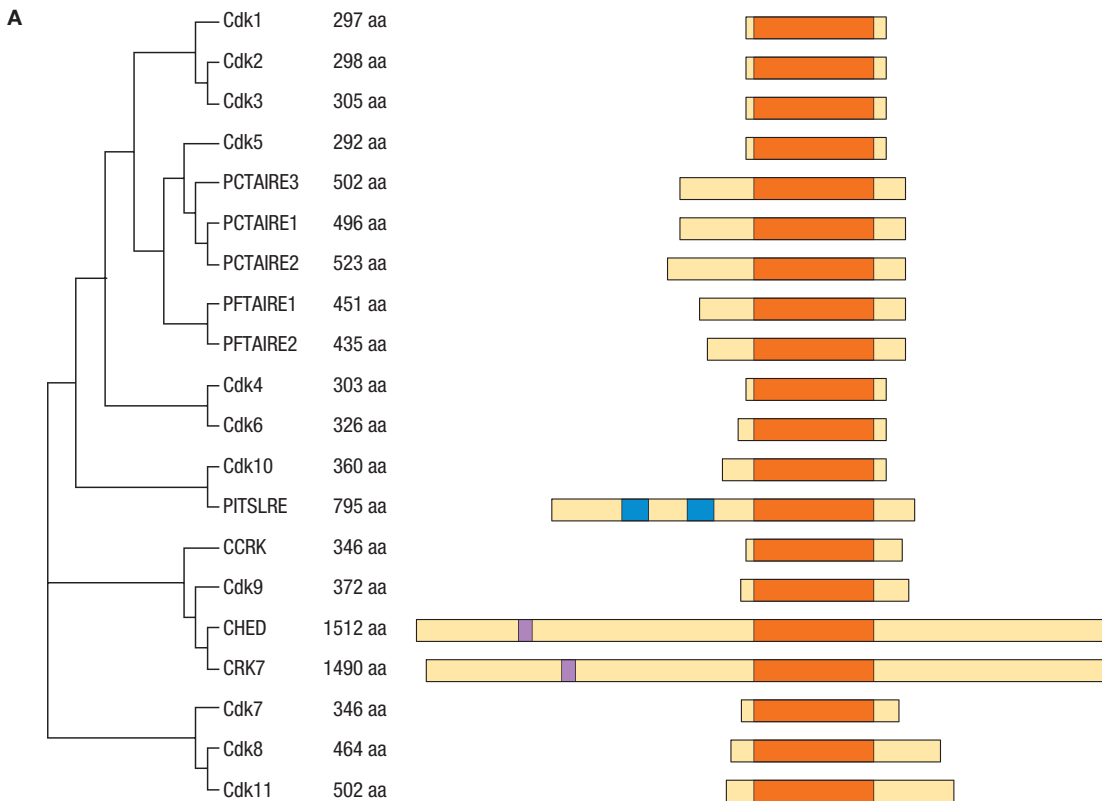


Figura 19.13 Proteine tirosina chinasi che agiscono come recettori transmembrana nell'uomo. L'acronimo del prototipo di ciascuna famiglia è mostrato nella parte in alto della figura; in basso sono indicati gli acronimi dei diversi membri delle famiglie. α e β indi-

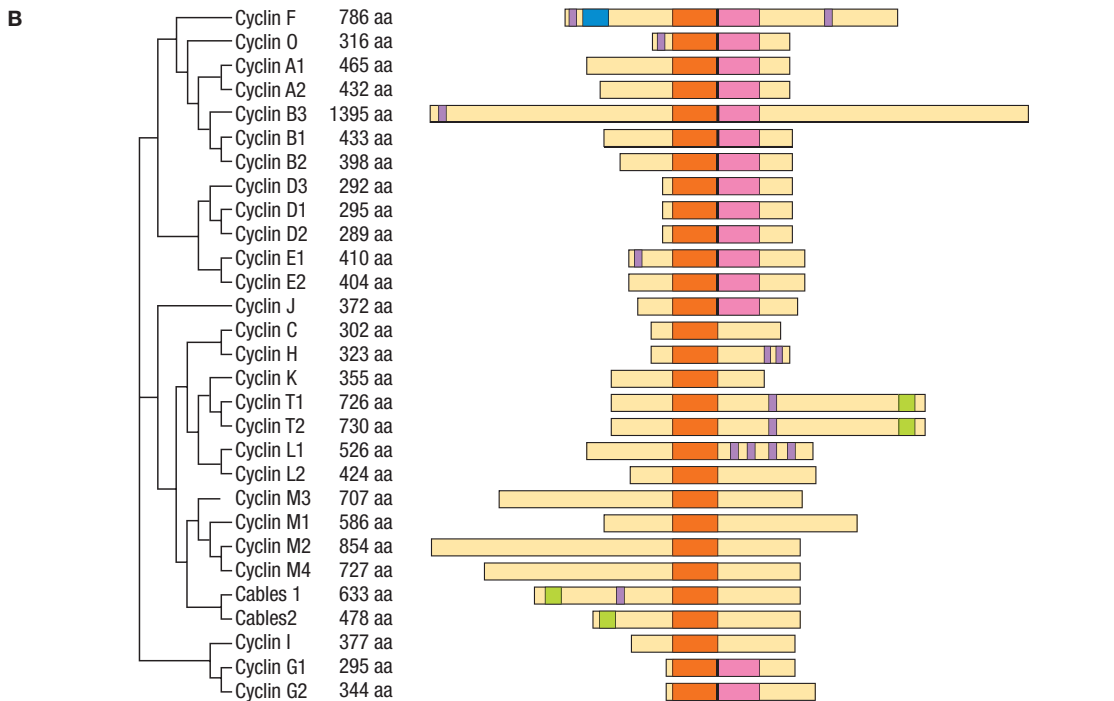
cano se lo stesso recettore è costituito da due subunità, mentre l'asterisco denota un membro della famiglia privo di attività chinasi intrinseca. I membri delle famiglie indicati in grassetto e in corsivo identificano geni implicati nell'insorgenza di tumori nell'uomo.

interazioni proteina-proteina tra chinasi e i loro substrati o i loro regolatori. Tali domini possono avere funzione inibitoria o attivatoria.

Riteniamo qui di fare un solo esempio che si collega a una famiglia di pS/pT proteine chinasi a cui abbiamo già accennato in altre sezioni del testo: le **CDK** o **chinasi ciclina-dipendenti**. Come descritto nei capitoli 6 e 7, tali chinasi costituiscono il motore del ciclo cellulare e controllano numerosi processi connessi con la proliferazione cellulare e vari aspetti del metabolismo del DNA. Le CDK sono state scoperte in studi pionieristici effettuati su eucarioti unicellulari, isolando mutanti alterati nella progressione del ciclo cellulare. Le CDK per funzionare efficientemente come chinasi devono associarsi a una subunità regolativa, nota come **ciclina**. Il numero di CDK e di cicline è via via aumentato nel corso dell'evoluzione: se nel lievito c'è una sola subunità catalitica e poche cicline, nell'uomo ne esistono molte di più (fig. 19.14). Diverse combinazioni tra CDK e cicline regolano passaggi critici del ciclo cellulare oltre ad altre funzioni importanti, quale la fosforilazione dell'RNA polimerasi II negli eucarioti (vedi cap. 11). Il legame con una ciclina aumenta di oltre 40 000 volte l'attività della subunità catalitica CDK: infatti, tale legame determina un riarrangiamento spaziale di alcuni amminoacidi critici nella CDK che è essenziale per il corretto legame con l'ATP e per il coordinamento di ioni Mg^{2+} richiesti per l'attività catalitica. Se le

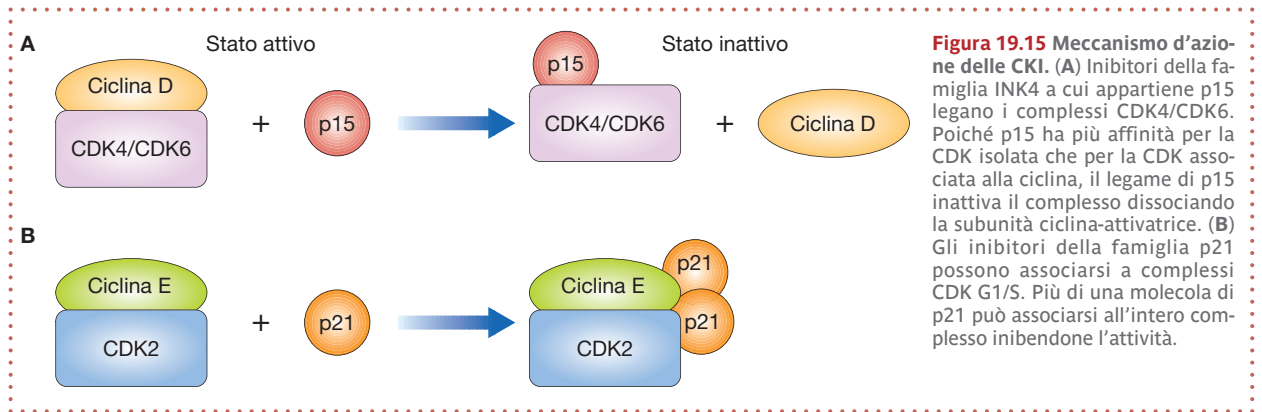


■ Dominio ser/thr proteine chinasi
 ■ Segnale di localizzazione nucleare
 ■ Dominio coiled-coil



■ Box ciclina N-terminale
 ■ Box ciclina C-terminale
 ■ F-box
 ■ Segnale di localizzazione nucleare
 ■ Box ricco in prolina

Figura 19.14 CDK e cicline nell'uomo. Vengono mostrati un elenco delle diverse CDK (A) e delle cicline (B) e una rappresentazione dei loro domini conservati (a sinistra).



cicline sono delle subunità con attività stimolatoria esistono altre proteine che interagiscono con le CDK inibendone l'attività e sono chiamate **Cyclin Kinase Inhibitors** o **CKI**. Esistono anche in questo caso famiglie di CKI che agiscono in modi diversi. Due esempi sono indicati in **figura 19.15**. La CKI nota come p15 appartiene alla famiglia denominata INK4: p15 inibisce specificamente i complessi CDK4 e CDK6 e la sua produzione è stimolata dalla citochina TGFβ. La CKI nota come p21, analogamente agli altri membri della famiglia p27 e p57, mostra un ampio spettro d'inibizione, inibendo diversi complessi CDK che agiscono nelle fasi G1/S. Contrariamente a p15, gli inibitori della famiglia p21 hanno un'affinità più alta per i complessi CDK-ciclina che per la CDK isolata e, pertanto, inibiscono l'attività dell'intero complesso. Le CDK sono considerate il "motore" del ciclo cellulare, mentre le CKI sono considerate i "freni". Mutazioni che inattivano le CKI portano, infatti, a un'incontrollata proliferazione cellulare e, quindi, all'insorgenza di tumori.

Poiché la fosforilazione di proteine ha un'importante funzione regolativa, possiamo prevedere che esistano diverse strategie per eliminare i gruppi fosfato dalle proteine fosforilate o per spegnere l'attività della chinasi. La prima strategia si può basare sull'utilizzo d'inibitori, come abbiamo descritto sopra per le CKI. La seconda è la distruzione proteolitica dell'eventuale subunità attivatrice: un altro modo con cui si inattivano le CDK. Le cicline, infatti, hanno una vita media molto breve e vengono rapidamente ubiquitinate e avviate alla via degradativa tramite il proteosoma. Questa strategia è frequentemente utilizzata nei meccanismi di controllo del ciclo cellulare, dove una proteina regolativa deve funzionare per un lasso di tempo limitato. Una terza via per spegnere una chinasi consiste nell'eliminare lo stimolo che è talvolta necessario per accenderne l'attività. Per esempio, l'attività di alcune chinasi dipende da piccole molecole di ligando, quali l'AMP ciclico o cAMP, per cui la loro attività può essere spenta eliminando il ligando stesso mediante l'azione di fosfodiesterasi che lo idrolizzano ad AMP.

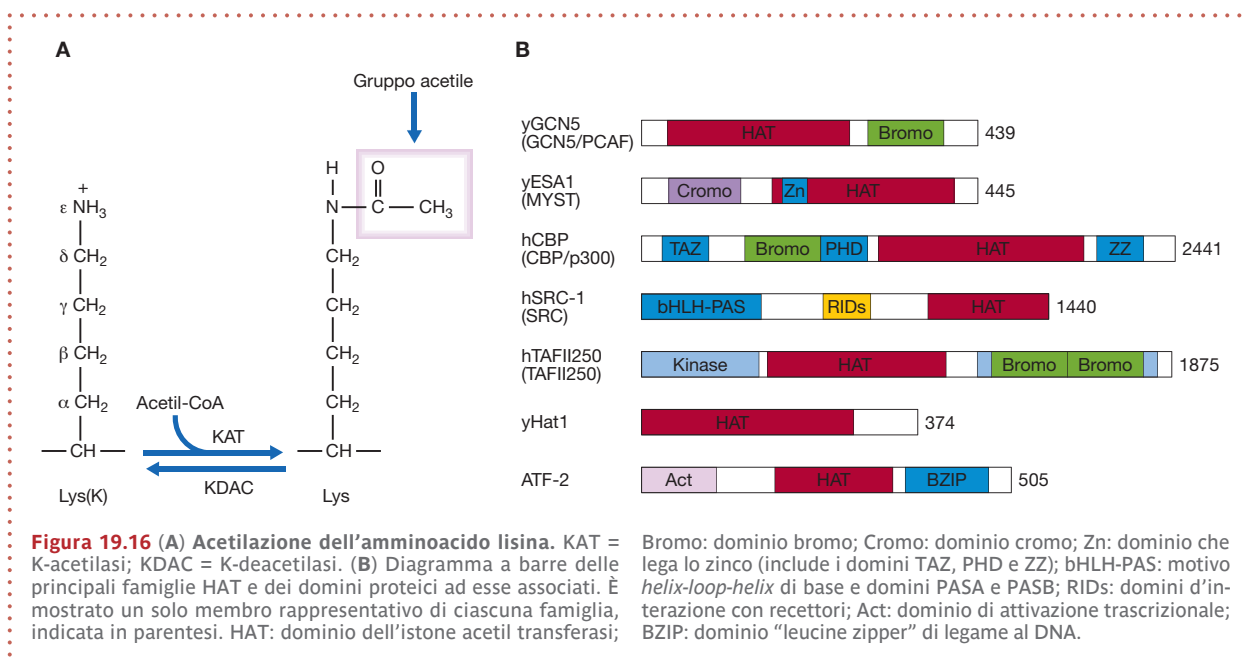
Il modo più diretto per eliminare gli effetti della fosforilazione di una proteina consiste, però, nell'idrolizzare direttamente il gruppo fosfato dai residui di tirosina, serina o treonina fosforilati. Esistono nel genoma umano circa **150 fosfatasi** in grado di catalizzare tali reazioni di defosforilazione. Come per le chinasi, una classe di fosfatasi defosforila residui di pS e pT, un'altra residui di pY e una terza classe tutti e tre i residui fosforilati. Mentre la specificità delle proteine chinasi nel fosforilare i propri substrati è molto elevata, la specificità delle fosfatasi è meno rigida dato che possono, almeno in parte, sostituirsi una con l'altra. Probabilmente per tale motivo, mutazioni nei geni codificanti per fosfatasi non sono così fortemente associate a patologie nell'uomo.

19.5 Acetilazione di proteine

L'**acetil-CoA** è il donatore di gruppi acetile che vengono covalentemente legati alle proteine sia all'estremità N-terminale che su residui interni di lisina. Il residuo amminoacidico N-terminale di oltre l'80% delle proteine eucariotiche è acetilato, mentre tale modificazione è estremamente rara nei procarioti. L'enzima che catalizza tale reazione si chiama **N-acetil transferasi (NAT)** e agisce co-traduzionalmente quando la proteina è lunga 20-50 amminoacidi ed è ancora in fase di sintesi sul ribosoma. La funzione biologica dell'acetilazione N-terminale di proteine non è chiara, ma mutanti difettivi in tale processo in lievito mostrano difetti fenotipici. Per esempio, la tropomiosina deve essere acetilata terminalmente per interagire efficientemente con l'actina.

L'interesse principale dell'**acetilazione** riguarda, però, la modificazione interna di proteine sul gruppo amminico ϵ di una lisina (**fig. 19.16A**). L'acetilazione degli istoni è particolarmente rilevante a tale riguardo e, come vedremo a breve, anche la loro metilazione. Infatti, tutti e quattro gli istoni H2A, H2B, H3 e H4 che formano il core ottamerico del nucleosoma sono sia acetilati che metilati e tali modificazioni sono concentrate nelle "code" N-terminali degli istoni, che, non essendo particolarmente coinvolte nella formazione del core nucleosomico, sono flessibili, poco strutturate e accessibili a modificazioni post-traduzionali. Abbiamo già ampiamente discusso (*vedi* capp. 4, 11 e 17) dell'importanza di tali modificazioni che rappresentano il cosiddetto **codice degli istoni**, che ha una grande rilevanza nei meccanismi di ereditarietà epigenetica, dato che tali modificazioni possono essere trasmesse di generazione in generazione. Bisogna anche sottolineare che le modificazioni istoniche sono tra loro connesse: per esempio, una certa modificazione su uno specifico residuo amminoacidico può avvenire solo se un altro residuo è stato precedentemente modificato.

L'interesse per l'acetilazione interna di proteine è via via aumentato quando si è scoperto che altre proteine coinvolte nella regolazione della struttura della cromatina e anche della trascrizione del DNA sono risultate essere acetilate. Non



solo, ma alcuni attivatori o repressori trascrizionali sono di fatto enzimi che realizzano tale modificazione (*vedi* cap. 11).

Il prototipo di fattore trascrizionale in grado di catalizzare l'acetilazione degli istoni (*Histone Acetyl Transferase* o **HAT**) è noto come **Gcn5**. È stato inizialmente scoperto in lievito, ma numerosi omologhi funzionali sono stati trovati in tutti gli eucarioti a costituire una famiglia di proteine. Domini HAT sono presenti in proteine che regolano la trascrizione e che hanno domini specifici multipli: ciò è in accordo con il ruolo multifunzionale di co-attivatori della trascrizione nell'influenzare interazioni proteina-proteina nell'assemblaggio di "macchine proteiche". Esistono diverse famiglie di HAT (**fig. 19.16B**) con domini funzionali distinti, quale il bromodominio che lega regioni proteiche contenenti lisine acetilate: il bromodominio può, quindi, essere coinvolto sia nell'interazione con gli istoni dopo che questi sono stati modificati dall'attività HAT sia nel modulare interazioni intramolecolari.

Non sono noti enzimi in grado di deacetilare la modificazione N-terminale di proteine. Al contrario, le lisine acetilate degli istoni sono deacetilate da enzimi noti come **HDAC** (*Histone Deacetylase*): queste proteine funzionano, per lo più, come corepressori della trascrizione (*vedi* cap. 11). Infatti, lo stato ipoacetilato delle lisine nelle code N-terminali degli istoni correla con la condensazione dei cromosomi, il silenziamento di promotori e, in genere, con l'assenza di trascrizione in quelle regioni del DNA. Ci sono 16 HDAC nel proteoma umano che si possono raggruppare in due famiglie: (1) nove HDAC classiche, che sono delle semplici idrolasi che rilasciano l'acetile come acetato; (2) sette sirtuine, che non sono semplici idrolasi, ma utilizzano il NAD come co-substrato rilasciando alla fine della reazione una molecola di lisina originariamente acetilata, una di nicotinammide e una molecola di 2'-O-acetil-ADP-ribosio. Non è chiaro perché coesistano questi due tipi di HDAC, dato che la deacetilazione è una reazione termodinamicamente favorita e le sirtuine, invece, utilizzano una molecola di NAD a ogni reazione. Le sirtuine sono state conservate nell'evoluzione dal lievito all'uomo, svolgono importanti funzioni per mantenere le code degli istoni H3 e H4 ipoacetilate e sorvegliano altri aspetti importanti nel controllo della stabilità del genoma e nei processi d'invecchiamento cellulare, soprattutto riguardanti la stabilità dei telomeri. È probabile che il NAD, un termometro del bilancio energetico, sia collegato meccanicisticamente alla reazione, forse modulando l'interazione con partner non ancora individuati.

19.6 Metilazione di proteine

La S-adenosilmetionina o SAM è il co-substrato donatore dei gruppi metile nella **metilazione di proteine** (**fig. 19.17**). La metilazione può avvenire su atomi d'azoto delle catene laterali degli amminoacidi lisina, istidina, arginina, glutammina e asparagina, su atomi d'ossigeno delle catene laterali carbossilate dell'a-

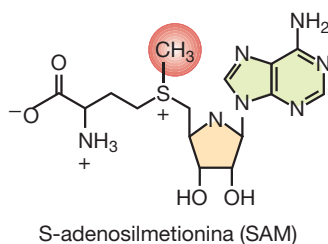


Figura 19.17 Formula chimica della S-adenosilmetionina (SAM).

cido glutammico e aspartico o sullo zolfo della catena laterale della cisteina. La metilazione sulle catene laterali carbossilate di aspartico e glutammico copre una carica negativa sostituendola con un gruppo idrofobico. La metilazione delle lisine non altera la carica cationica ma aggiunge idrofobicità. In particolare, se una lisina viene di- o trimetilata non solo si aumenta l'idrofobicità, ma si incrementa notevolmente anche l'ingombro sterico, il che può influire notevolmente sulle interazioni proteina-proteina.

La metilazione di amminoacidi è frequente sia in procarioti che in eucarioti e modula la funzionalità di proteine in modi molto diversi. Qui, però, discuteremo brevemente tale modificazione post-traduzionale solo nell'ambito della sua importanza nel modulare fenomeni connessi al controllo dell'espressione genica.

Ancora una volta, il bersaglio preferito per la metilazione sono le code N-terminali degli istoni H3 e H4. L'istone H3 è quello più modificato post-traduzionalmente: alcune lisine possono essere sia metilate che acetilate; altre lisine possono essere solo acetilate o metilate; arginine possono essere metilate e, anche se non ne abbiamo discusso, alcune serine sono fosforilate. Considerando tutte le possibili modificazioni, il numero di isoforme proteiche diverse per una molecola di istone H3 è enorme: 110 592 possibili combinazioni di modificazioni post-traduzionali! Per di più, gli amminoacidi possono essere mono-, di- o trimetilati e va considerato che tutte queste modificazioni possono essere teoricamente diverse nei circa 10^7 nucleosomi presenti sul genoma umano, ciascuno contenente due copie dell'istone H3 (*vedi* cap. 4). Il numero di possibili modificazioni nel genoma diventa quindi astronomico! Esistono **metil transferasi** specifiche per i diversi residui di lisina e arginina negli istoni H3 e H4. In **figura 19.18** è mostrato uno schema semplificato che indica soltanto le metil transferasi che metilano residui specifici di lisina nelle code N-terminali degli istoni H3 e H4. Altre ne esistono che modificano amminoacidi nelle regioni degli istoni che fanno parte del core del nucleosoma.

Solo più recentemente sono stati scoperti gli enzimi in grado di demetilare i diversi residui modificati degli istoni, e anch'essi sono indicati in **figura 19.18**. Sorprendentemente è stato scoperto che metilasi e **demetilasi** possono fare parte dello stesso complesso multiproteico (**fig. 19.19**). La recente scoperta delle demetilasi e il fatto che esse possono co-esistere con le metilasi hanno reso molto più dinamica la visione della struttura della cromatina. Infatti, la metilazione degli istoni è il risultato di un equilibrio tra attività enzimatiche che hanno effetti opposti, probabilmente controllati da regolatori trascrizionali sequenza-specifici.

19.7 Glicosilazione e modificazioni lipidiche di proteine

Concludiamo questa breve descrizione delle modificazioni post-traduzionali di proteine accennando alle modificazioni di proteine tramite l'aggiunta di zuccheri e di acidi grassi. Non potevamo tralasciare completamente tali modificazioni, dato che hanno funzioni molto rilevanti nella biologia della cellula.

La **glicosilazione** è un complesso sistema di modificazione che può avvenire sia co- che post-traduzionalmente e, come tutte le modificazioni di proteine, ha la funzione di espandere la diversità del proteoma. La glicosilazione avviene sia in procarioti che archeobatteri ed eucarioti, anche se con modalità diverse. Negli eucarioti è noto che la glicosilazione di proteine gioca un ruolo molto importante in una varietà di processi biologici: la risposta immunitaria, il riconoscimento tra cellule, la localizzazione intracellulare, la stabilità e la struttura di proteine; nei procarioti è legata alla loro patogenicità e all'invasione dell'ospite.

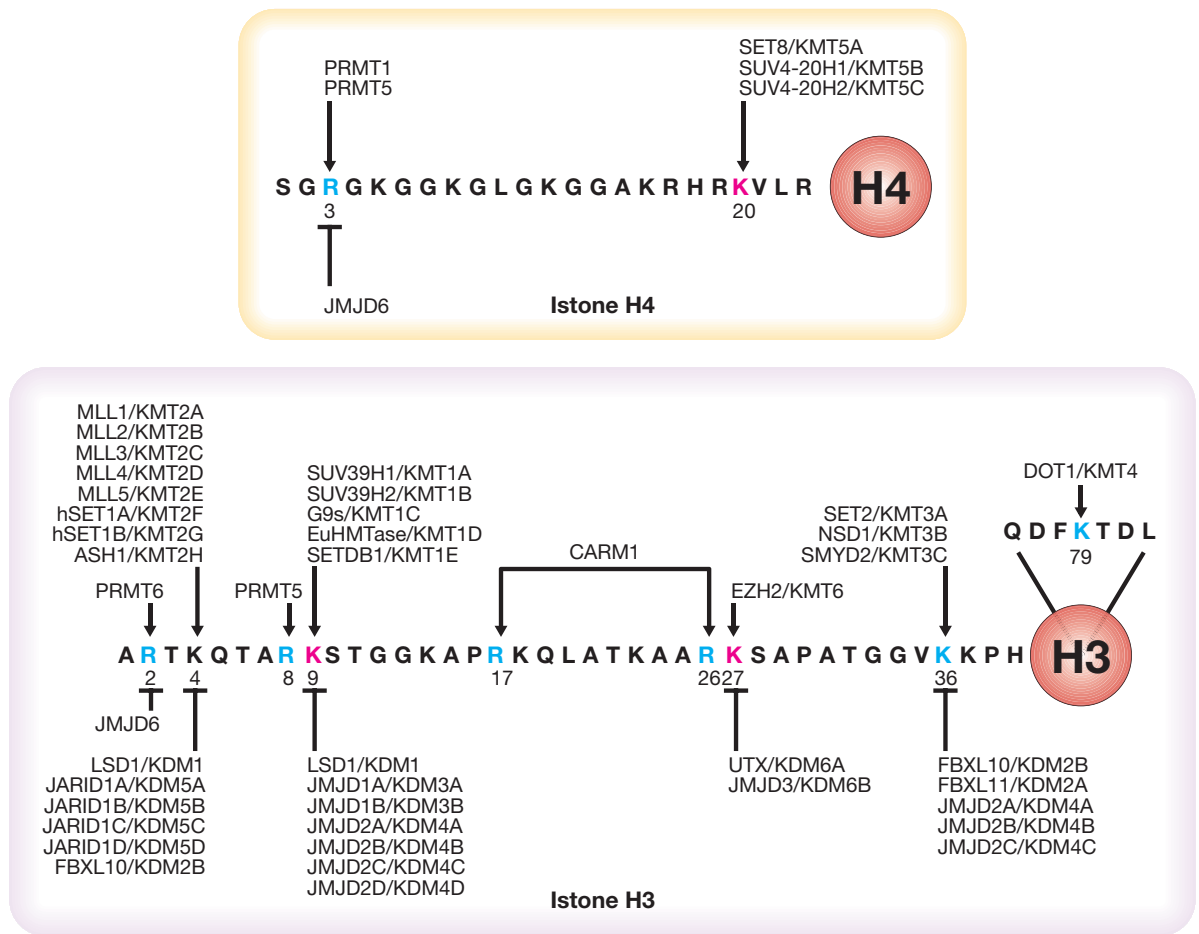


Figura 19.18 Schema delle code N-terminali degli istoni H3 e H4. Gli amminoacidi arginina (R) o lisina (K) che sono metilati sono indicati in azzurro o in rosso se tale modificazione correla, rispettivamente, con l'attivazione o la repressione della trascrizione. Le metil transferasi istoniche responsabili delle modificazioni

sono indicate sopra la sequenza, mentre le corrispondenti demetilasi sono indicate sotto la sequenza. Due nomi sono indicati per ciascun enzima: il nome originario e quello proposto recentemente per ottenere una classificazione più chiara.

Esistono sostanzialmente due tipi diversi di glicosilazione, la N-glicosilazione e la O-glicosilazione, che differiscono per vari aspetti: meccanismo d'azione, compartimento cellulare in cui vengono realizzate e sito d'attacco delle catene glicosidiche.

La **N-glicosilazione** si realizza tramite l'aggiunta di una catena glucidica all'atomo di azoto di una catena laterale di asparagina e, negli eucarioti, ha inizio nel reticolo endoplasmatico quando la proteina è ancora in fase di sintesi. Il primo evento consiste nel legame di una catena di 14 zuccheri (2 molecole di N-acetilglucosammina, 3 di glucosio e 9 di mannosio) a un residuo laterale di asparagina. L'oligosaccaride coinvolto nella modificazione è assemblato anch'esso nel reticolo endoplasmatico ed è trasferito come un unico elemento da una molecola di dolico fosfato alla proteina tramite l'azione dell'enzima glicosil transferasi. L'attacco della catena glucidica avviene di solito su asparagine che si trovano in contesti amminoacidici specifici.

Dopo l'attacco, la catena glucidica può essere ulteriormente modificata eliminando alcuni residui di zuccheri. Le proteine che hanno subito l'attacco della catena glucidica e la sua successiva modificazione sono poi trasportate all'apparato

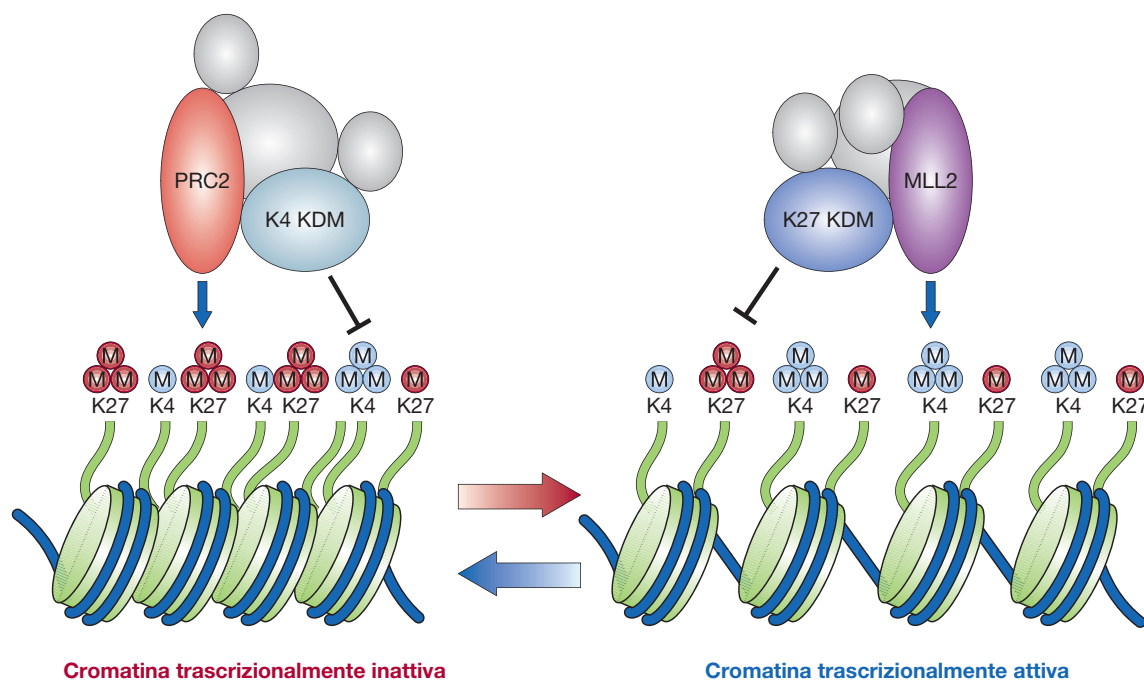


Figura 19.19 Modello di cooperazione tra metil transferasi e demetilasi che appartengono a complessi diversi e agiscono su alcuni specifici residui di lisina. La presenza di due attività enzimatiche opposte in entrambi i complessi permette la loro cooperazione nel controllo trascrizionale. Questi complessi sono sempre attivi e un certo stato di metilazione degli istoni è il

risultato dell'equilibrio tra le attività dei due complessi. Qui sono mostrati i due complessi noti come MLL2 e PRC2, che sono le metilasi, mentre K27 KDM e K4 KDM sono le demetilasi. Sono indicati i residui di lisina (K) sulla coda dell'istone H3 e lo stato di metilazione (mono o tri). Lo stato attivo della cromatina è mostrato a destra.

di Golgi. Qui subiscono una serie di ulteriori modificazioni che sono specifiche per ciascuna proteina.

La **O-glicosilazione** è un processo altamente specifico che non prevede le modificazioni seriali della proteina descritte per la N-glicosilazione nel reticolo endoplasmatico. Si svolge, infatti, nell'apparato di Golgi, dove vari zuccheri possono legarsi alla proteina a livello dell'ossigeno delle catene laterali di residui di serina e treonina. Viene aggiunto uno zucchero alla volta e, solitamente, il numero di molecole di zucchero aggiunte nel processo è limitato a pochi residui.

Un numero notevole di proteine subisce anche una serie di modificazioni covalenti che prevedono l'aggiunta di una o più molecole di lipidi che indirizzano la proteina modificata su varie membrane, quali il reticolo endoplasmatico, l'apparato di Golgi, i lisosomi, le membrane cellulari, i dendriti, gli assoni ecc. Le **modificazioni lipidiche** più frequenti sono: (1) la miristoilazione dell'N-terminale di un residuo di glicina; (2) la palmitoilazione, farnesilazione o geranilazione dell'atomo di zolfo della catena laterale di una cisteina; (3) l'attacco sul C-terminale di una proteina di un'ancora di GPI (una molecola complessa formata da fosfatidilinositolo e oligosaccaridi).

Tramite quest'ultima modificazione la proteina si ancora a una membrana, come esemplificato in **figura 19.20**. Proteine modificate tramite legami con lipidi svolgono funzioni cellulari molto importanti, quali proteine G accoppiate a recettori, proteine che costituiscono i canali ionici delle membrane, proteine chinasi, proteine delle sinapsi, proteine dei capsidi virali.

In questa breve panoramica delle modificazioni post-traduzionali di proteine ne abbiamo tralasciate numerose anche importanti, quali ADP-ribosilazione,

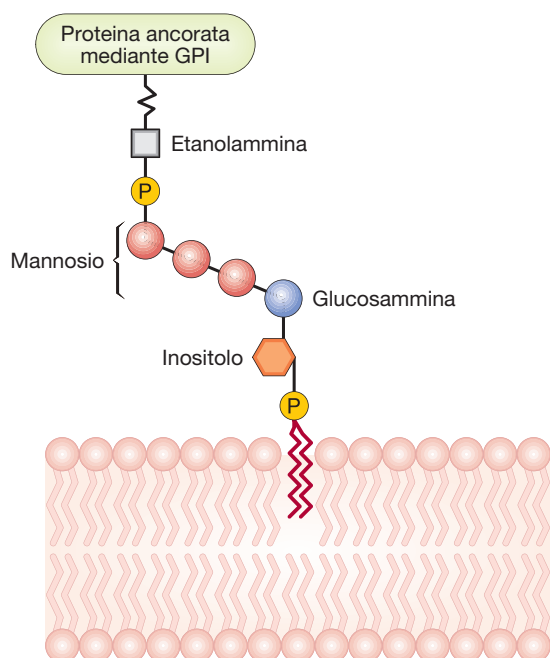


Figura 19.20 Un'ancora GPI connette il C-terminale di una proteina a una membrana tramite un aggancio che coinvolge una porzione oligosaccaridica e un fosfolipide. Molte proteine negli eucarioti sono legate alla membrana cellulare mediante un'ancora GPI (glicosil fosfatidilinositolo) i cui componenti sono indicati in figura. L'ancora GPI è assemblata nel reticolo endoplasmatico e in tale sede viene legata alla proteina. Successivamente l'intero complesso proteina-GPI è trasportato sulla superficie cellulare.

modificazioni tramite ossido-riduzioni, carbossilazione, amidazione ecc. Come già accennato, abbiamo scelto di trattare brevemente solo le modificazioni post-traduzionali che più sono connesse a fenomeni di regolazione dell'espressione genica, che è l'argomento che caratterizza tutto il testo.

Lecture di approfondimento

- Agger *et al.* (2008), «The emerging functions of hystone demethylases», *Current Opinion in Genetics and Development*, **18**, pp. 159-168.
- Alonso, A. *et al.*, (2004), «Protein tyrosine phosphatases in the human genome», *Cell*, **117**, pp. 699-711.
- Finley, D. (2009), «Recognition and processing of ubiquitin-protein conjugates by the proteasome», *Annual Review of Biochemistry*, **78**, pp. 477-513.
- Geiss-Friedlander, R. e Melchior F. (2007), «Concepts in sumoylation: a decade on», *Nature Reviews Molecular Cell Biology*, **8**, pp. 947-956.
- Ubersax, J.A. e Ferrell, Jr. J.E. (2007), «Mechanisms of specificity in protein phosphorylation», *Nature Reviews Molecular Cell Biology*, **8**, pp. 530-542.



APPROCCI, TECNICHE E MODELLI IN BIOLOGIA MOLECOLARE

Spesso la Biologia Molecolare viene confusa con un insieme di tecniche di laboratorio. Ciò costituisce una rappresentazione sbagliata della disciplina, dato che non basta saper applicare una tecnologia, anche estremamente utile e diffusa come per esempio la PCR, per potersi considerare un biologo molecolare. In realtà, tale disciplina ha l'ambizione di voler comprendere a livello molecolare fenomeni biologici complessi e, inevitabilmente, per raggiungere tale traguardo ha dovuto sviluppare tecniche via via sempre più complesse e sofisticate. Di fatto, nei circa cinquant'anni in cui la Biologia Molecolare è cresciuta, è stato proprio lo sviluppo di particolari tecnologie a permettere accelerazioni imprevedibili nelle nostre conoscenze biologiche.

Nel capitolo 20 descriveremo solo le più importanti tra le numerosissime tecniche sviluppate dai biologi molecolari che hanno permesso un impressionante miglioramento delle conoscenze sulla struttura e funzione di macromolecole biologiche, quali acidi nucleici e proteine.

Il capitolo 21 è dedicato alla Bioinformatica. Il sequenziamento di un gran numero d'interi genomi e la conoscenza sempre più approfondita della struttura tridimensionale di numerose proteine hanno prodotto una quantità di informazioni che sarebbe rimasta per lo più inutilizzata senza l'ausilio della Bioinformatica, una nuova disciplina che si è sviluppata in parallelo con la Biologia Molecolare. Gli obiettivi principali della Bioinformatica sono: (1) lo sviluppo degli strumenti più appropriati per organizzare banche dati e i relativi sistemi d'interrogazione; (2) lo sviluppo di metodologie innovative per l'analisi dei dati che possano suggerire possibili funzioni biologiche, facilitando la successiva e necessaria verifica sperimentale.

La Biologia Molecolare ha certamente tra i suoi obiettivi un perfezionamento delle conoscenze dei processi biomolecolari più importanti al fine di migliorare le condizioni di vita e la salute dell'uomo. Tuttavia, per ovvi motivi etici e per le loro proprietà biologiche, esseri umani o anche sistemi cellulari di mammifero

in vitro spesso non sono il sistema sperimentale migliore per analizzare fenomeni biologici complessi a livello molecolare. Pertanto, la Biologia Molecolare utilizza per i propri studi degli organismi modello che permettono l'acquisizione di conoscenze sofisticate che possono poi essere estrapolate all'uomo e verificate in cellule umane. Ciò è possibile perché la maggior parte dei meccanismi molecolari che regolano la vita di una cellula e di un organismo sono stati estremamente conservati nel corso dell'evoluzione. Proprietà e caratteristiche dei più importanti sistemi biologici modello verranno discusse nel capitolo 22.

20

Tecniche della Biologia Molecolare

- 20.1 Tecniche spettrofotometriche per l'analisi della denaturazione e riassociazione del DNA
- 20.2 Uso di sonde di DNA per l'identificazione e analisi di sequenze nucleotidiche
- 20.3 Ultracentrifugazione
- 20.4 Tecnologie di base per l'isolamento e la manipolazione di geni
- 20.5 Banche di DNA
- 20.6 PCR: reazione a catena della polimerasi
- 20.7 Determinazione della sequenza nucleotidica del DNA
- 20.8 Piattaforme di sequenziamento di nuova generazione
- 20.9 Metodologie recenti per la modificazione mirata dei genomi ("genome editing")
- 20.10 Valutazione dell'espressione di un gene
- 20.11 Inibizione dell'espressione di geni specifici: RNA antisenso e interferenza da RNA
- 20.12 Interazioni tra macromolecole biologiche

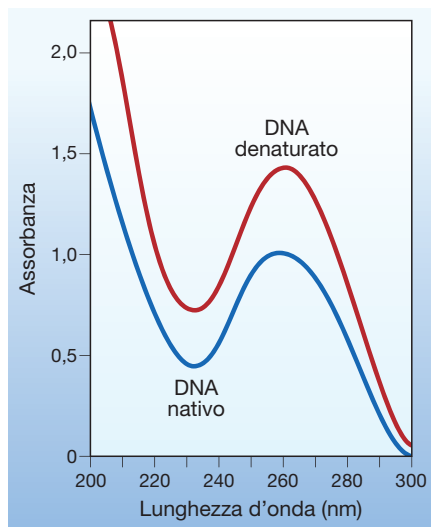
20.1 Tecniche spettrofotometriche per l'analisi della denaturazione e riassociazione del DNA

Alcune delle tecniche che descriveremo nel capitolo possono sembrare superate in quanto sostituite, negli anni recenti, dall'analisi diretta delle sequenze nucleotidiche. Tuttavia, riteniamo che sia importante descriverne alcune, quali la misura della temperatura di fusione, l'analisi della cinetica di riassociazione del DNA e l'ibridazione di sonde specifiche, perché, anche se non più usate come alcuni anni fa, sono alla base di molte tecniche più moderne.

Spettro di assorbimento

Come abbiamo discusso nel paragrafo 2.3, i due filamenti di DNA sono tenuti insieme nella struttura a doppia elica da un gran numero di legami non covalenti deboli: legami idrogeno tra le basi complementari e interazioni idrofobiche dovute all'impilamento delle basi. Questi legami possono essere indeboliti da vari agenti, quali alta temperatura, alto pH e alcuni agenti chimici (ad es. urea), fino a portare alla separazione dei due filamenti. Questo processo si chiama **denaturazione del DNA** e, come vedremo più avanti, è un processo reversibile. Per studiare la denaturazione del DNA sono stati messi a punto metodi che permettono di seguire la denaturazione del DNA in soluzione misurando l'assorbimento della luce

Figura 20.1 Spettro di assorbimento del DNA. Una soluzione di DNA alla concentrazione di 50 $\mu\text{g/mL}$ è stata posta nella cuvetta di uno spettrofotometro e ne è stato determinato lo spettro di assorbimento nell'intervallo di lunghezza d'onda tra 200 e 300 nm, cioè nella regione della luce ultravioletta (curva blu). Lo spettro che si ottiene per il DNA, ma anche per l'RNA, presenta un tipico picco alla lunghezza d'onda di 260 nm. Per questa soluzione alla concentrazione di 50 $\mu\text{g/mL}$, la lettura di assorbanza (o densità ottica) a 260 nm è di 1,0. La curva rossa mostra lo spettro ottenuto per la stessa soluzione di DNA alla concentrazione di 50 $\mu\text{g/mL}$ in cui però i due filamenti della doppia elica sono stati separati per denaturazione. In questo caso lo spettro mostra lo stesso andamento e lo stesso picco a 260 nm, ma l'assorbanza aumenta di circa il 40%.



ultravioletta con uno spettrofotometro. Il DNA, come pure l'RNA, non assorbe radiazioni elettromagnetiche nell'intervallo di lunghezza d'onda che corrisponde alla luce visibile; infatti, soluzioni di DNA e RNA sono al nostro occhio perfettamente trasparenti. Come mostrato per il DNA in **figura 20.1**, tali soluzioni assorbono però luce ultravioletta con un caratteristico picco a 260 nm. La misura con uno spettrofotometro dell'**assorbanza A** (talvolta chiamata anche **OD = Densità Ottica** od *Optical Density*) a questa lunghezza d'onda viene comunemente usata in laboratorio per determinare la concentrazione del DNA (o dell'RNA) in soluzione. Infatti, è stato determinato che una lettura di A pari a 1,0 corrisponde a circa 50 $\mu\text{g/mL}$ di DNA nativo a doppio filamento. Ma, come anche mostrato in **figura 20.1**, l'assorbanza aumenta di circa il 40% quando il DNA viene denaturato. Infatti, l'assorbimento a 260 nm è dovuto alle basi puriniche e pirimidiniche e queste sono più mascherate nella struttura a doppia elica che non nei singoli filamenti del DNA denaturato. Questo fenomeno è chiamato **effetto ipercromico**.

Denaturazione del DNA, T_m

L'effetto ipercromico ci dà la possibilità di seguire la denaturazione del DNA utilizzando uno spettrofotometro in cui è possibile variare la temperatura della soluzione di DNA mentre, contemporaneamente, ne misuriamo l'assorbanza alla lunghezza d'onda di 260 nm. La curva rossa in **figura 20.2** si riferisce a un tipico esperimento di denaturazione di un DNA naturale. La temperatura a cui è avvenuta metà transizione ipercromica, e quindi metà denaturazione, viene chiamata T_m (*melting temperature* = **temperatura di fusione**).

La T_m dipende da vari parametri legati al tipo di soluzione in cui facciamo avvenire la denaturazione: infatti, una bassa forza ionica e/o un alto pH della soluzione acquosa in cui si trova il DNA destabilizzano la doppia elica; al contrario, un'alta forza ionica e un pH neutro la stabilizzano. È per questo che, in genere, per potere confrontare la T_m di diverse preparazioni di DNA e per potere rendere comparabili i risultati ottenuti in diversi laboratori, si esegue la misura della T_m in condizioni standardizzate (NaCl 0,15 M, Na citrato 0,015 M, pH 7,0). È stato così osservato che, a parità di condizioni di forza ionica, differenti preparazioni di DNA con composizione in basi diversa presentano curve di denaturazione diverse. In particolare, DNA ricchi in coppie di basi C-G, la cui interazione comporta tre legami idrogeno, presentano una T_m più alta rispetto a DNA con alto contenuto

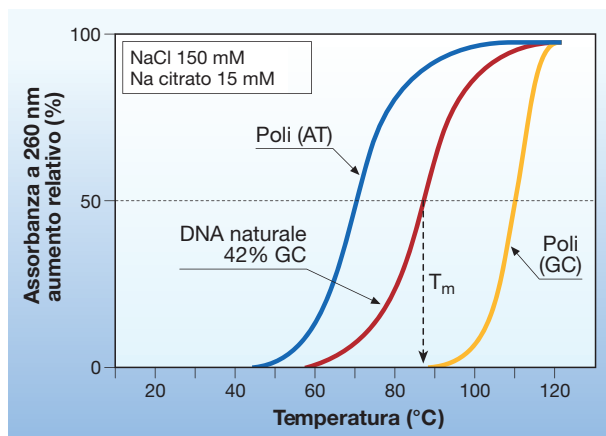


Figura 20.2 Denaturazione del DNA, T_m . Durante la denaturazione del DNA, la transizione da una conformazione a doppia elica a una conformazione a singolo filamento è accompagnata da un aumento dell'assorbanza (effetto ipercromico, vedi fig. 20.1). In rosso è riportata la curva di denaturazione di un DNA naturale: la T_m (T_m = melting temperature, temperatura di fusione) corrisponde alla temperatura a cui metà delle molecole sono state denaturate, cioè a circa 87 °C per il DNA naturale. In blu e giallo sono riportate le curve di denaturazione di duplex formati rispettivamente di sole coppie di basi A-T e di sole coppie di basi G-C.

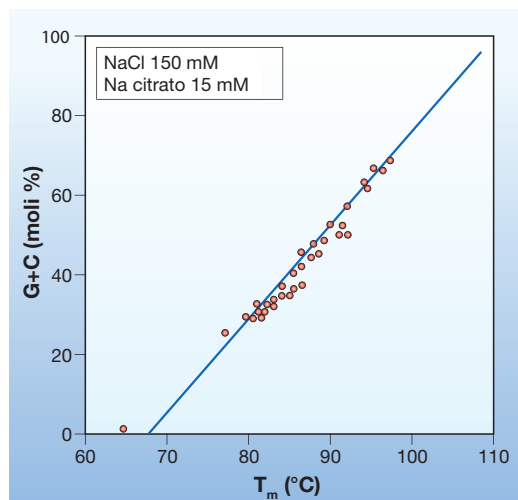


Figura 20.3 Dipendenza della temperatura di fusione (T_m) dal contenuto di G-C nel DNA. La misura della T_m di preparazioni di DNA ottenute da diversi virus, plasmidi e cellule, è stata fatta in condizioni standardizzate di forza ionica e pH: 0,15 M NaCl, 0,015 M Na citrato, pH 7,0.

di coppie A-T, che sono legate da solo due legami idrogeno. Come esempio, la stessa **figura 20.2** mostra le curve di denaturazione di due casi estremi: un DNA sintetico costituito di sole coppie di basi A-T, che mostra una T_m di oltre 15 °C più bassa del DNA naturale, e un DNA sintetico costituito di sole coppie di basi C-G, che ha una T_m di oltre 20 °C superiore al DNA naturale.

Come mostrato in **figura 20.3**, l'analisi della T_m di DNA purificati da numerosi organismi diversi ha messo in evidenza una relazione lineare tra T_m e contenuto in G-C del DNA, tanto che, per molti anni, la misura della T_m è stata uno dei due metodi utilizzati per determinare la composizione in basi del DNA (l'altro metodo è la misura della densità mediante ultracentrifugazione in gradiente di CsCl di cui parleremo più avanti).

Riassociazione del DNA, cinetica di riassociazione, Cot

La denaturazione del DNA è un processo reversibile. Infatti, se messi in opportune condizioni di temperatura e forza ionica, i due filamenti complementari separati per denaturazione possono riformare una doppia elica (duplex). In questo caso si parla di **rinaturazione**, o meglio di **riassociazione** dei filamenti di DNA (in inglese viene usato anche il termine *annealing*). Analogamente, si può formare una struttura a doppia elica anche tra una molecola di DNA e una di RNA, purché

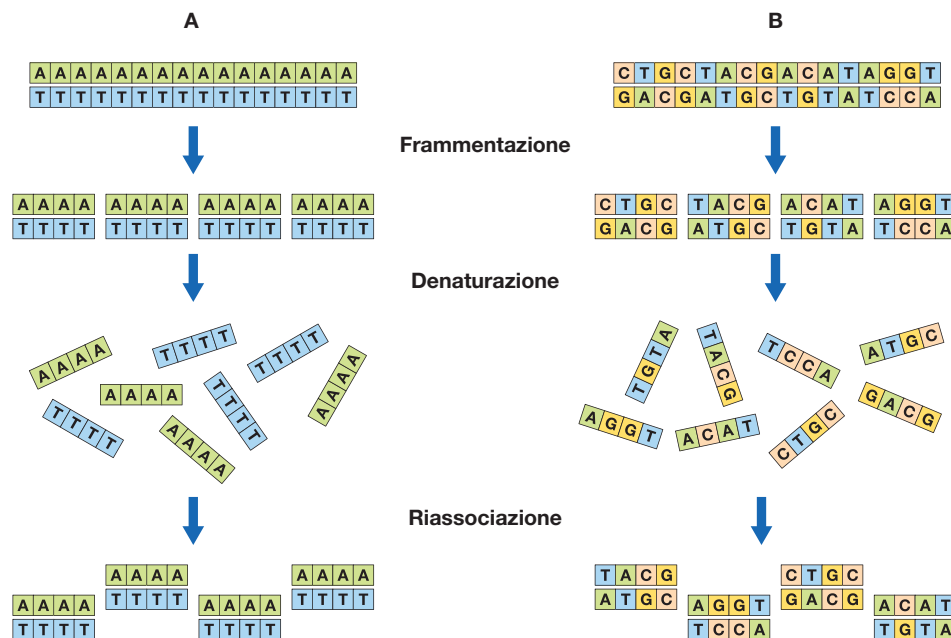


Figura 20.4 Rappresentazione semplificata e schematica della riassociazione del DNA. Sequenze semplici riassociano più rapidamente di sequenze complesse. (A) DNA con sequenza molto semplice in cui un filamento è costituito di sole A e l'altro, complementare, di sole T (complessità 1). (B) DNA un po' più complesso costituito da una sequenza eterogenea di 16 coppie di basi (complessità 16). Ambedue i DNA vengono frammentati alla lunghezza di 4 pb (in questo esempio semplificato, in realtà alla lunghezza di

300-400 pb) e posti in condizione di riassociare. La possibilità di un filamento di incontrarne uno complementare è maggiore per il campione A rispetto al campione B. Infatti, ogni singolo filamento di A ha il 50% delle probabilità di incontrarne uno complementare con il quale può riassociare e il 50% di incontrarne uno uguale a se stesso con il quale non può riassociare; invece ogni singolo filamento del DNA B ha solo il 12,5% delle probabilità di incontrare il complementare.

abbiano sequenze nucleotidiche complementari; in questo caso si parla di **ibridazione**. Va detto, però, che il termine "ibridazione" viene spesso usato anche per indicare l'associazione tra due filamenti di DNA quando essi non siano derivati per denaturazione di una doppia elica ma siano di origine diversa, come il caso, per esempio, dell'associazione tra una sequenza sonda e una sequenza bersaglio.

Le "opportune" condizioni da adottare in un esperimento di riassociazione o di ibridazione sono una combinazione di temperatura e forza ionica tali da trovarsi circa 25 °C sotto la T_m della doppia elica che andiamo a ricostituire. Tale scelta è basata su varie considerazioni, ma soprattutto sull'osservazione che la temperatura adottata influenza molto la specificità della riassociazione. Infatti, a temperature relativamente alte e abbastanza vicine alla T_m , avremo una riassociazione non molto efficiente ma sicuramente molto "stringente", cioè si appaieranno solo filamenti perfettamente complementari o quasi. Al contrario, a temperature abbastanza al di sotto della T_m si avrà una riassociazione più efficiente ma meno stringente, saranno cioè stabili anche interazioni tra filamenti non perfettamente complementari. Come abbiamo detto, in genere si considera che 25 °C al di sotto della T_m sia un buon compromesso tra queste varie esigenze. In alcuni casi si può preferire di utilizzare condizioni diverse a seconda se vogliamo privilegiare la stringenza (specificità) della reazione o la sua efficienza.

Messe nella condizione di riassociare, le molecole di DNA a singolo filamento, mosse da agitazione termica, avranno numerosi contatti casuali fino a quando si incontreranno due regioni perfettamente complementari e con polarità 5'-3' opposta. Queste formeranno i canonici legami idrogeno tra le basi e assumeranno la struttura a doppia elica, stabilizzata anche dalle interazioni idrofobiche dovute all'impilamento delle basi. Una volta avvenuto un evento di

appaiamento produttivo (**nucleazione**), se i filamenti sono veramente complementari la riassociazione proseguirà molto rapidamente lungo tutta la molecola. A parità di lunghezza delle molecole e di condizioni di incubazione (temperatura e forza ionica), la riassociazione di DNA denaturato è una reazione di secondo ordine che dipende dall'incontro delle due molecole complementari, che, per definizione, sono presenti a uguale concentrazione. La velocità con cui procede la riassociazione ci può dare informazioni sulla complessità del DNA che stiamo analizzando. Infatti, a parità di concentrazione del DNA (ad es. 10 mg/mL), un DNA con sequenza semplice riassocerà più rapidamente di un DNA con sequenza complessa, perché ciascun filamento ha una probabilità più alta d'incontrare nella soluzione un filamento con sequenza complementare. Questo concetto è spiegato in maniera molto schematica in **figura 20.4**. Va notato che, poiché quello che qui ci interessa è la velocità di riassociazione come misura della probabilità di incontro tra filamenti complementari, è opportuno che il DNA da analizzare venga prima frammentato a una misura standard di circa 300-400 pb. In questo modo la misura della riassociazione rifletterà soprattutto gli eventi di nucleazione.

La **figura 20.5** mostra le curve di riassociazione di DNA di diversa origine. Poiché alla probabilità di incontro dei filamenti complementari contribuiscono, linearmente, sia il tempo (t) della reazione che la concentrazione (C_0) dei filamenti di DNA, il parametro che definisce la velocità di riassociazione, e che è riportato sull'asse delle ascisse, è il prodotto: concentrazione (C_0) $\times$ tempo (t). Esso viene chiamato **Cot**, la cui unità di misura è moli $\times$ sec/L. Il valore di Cot a cui un DNA raggiunge metà riassociazione viene chiamato **Cot $_{1/2}$** . Va anche notato che, poiché i valori di Cot che ci interessano si estendono su 6 o 7 ordini di grandezza, essi sono riportati sull'asse delle ascisse in scala logaritmica. In ordinata abbiamo la percentuale di DNA denaturato, che quindi parte al 100% all'inizio dell'esperimento di riassociazione per terminare, a fine reazione, a valori vicini allo zero. Quattro delle cinque curve mostrate in **figura 20.5** si riferiscono a genomi di diversa complessità che non contengono quantità apprezzabili di sequenze ripetute e, quindi, si presentano come semplici curve di riassociazione sigmoidali. I Cot $_{1/2}$ di queste curve (intersezioni delle curve con la retta al 50% di riassociazione) si differenziano in maniera direttamente proporzionale alle differenze di complessità dei genomi. Per esempio, il Cot $_{1/2}$ del DNA del fago MS2 (lunghezza del genoma $3,5 \times 10^3$ pb) è di circa $3,5 \times 10^3$ volte maggiore del Cot $_{1/2}$ del DNA sintetico poli(A):poli(T) (complessità 1 pb), e il Cot $_{1/2}$ del DNA di *Escherichia coli* (lunghezza del genoma circa $4,2 \times 10^6$ pb) è di circa 25 volte

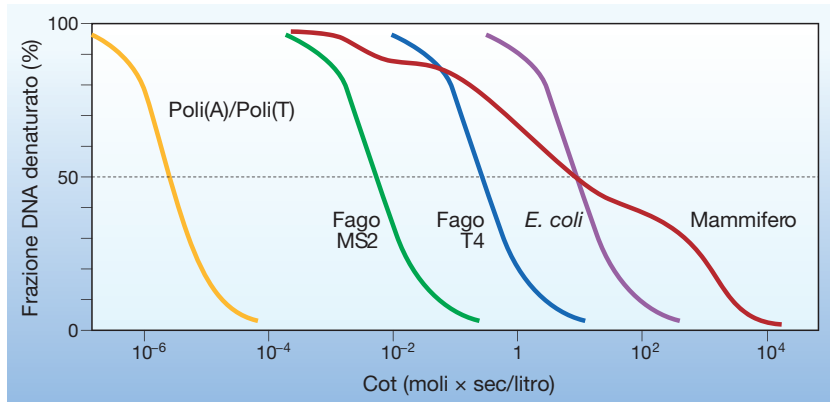


Figura 20.5 Cinetica di riassociazione di diversi DNA. La cinetica di riassociazione è stata determinata per DNA con diverse complessità. Curva gialla: poli(A)/poli(T), complessità 1 pb; curva verde: fago MS2, complessità ≈ 3500 pb; curva blu: fago T4, complessità $\approx 170\,000$ pb; curva viola: *E. coli*, complessità $\approx 4,2 \times 10^6$ pb. Come si può osservare, il Cot $_{1/2}$ è inversamente proporzionale alla grandezza del genoma. La curva rossa descrive la cinetica di riassociazione del DNA genomico dei mammiferi, in cui si osserva la presenza di tre componenti che riassociano con cinetiche diverse: DNA altamente ripetitivo, DNA mediamente ripetitivo e DNA unico.

maggiore del $Cot_{1/2}$ del DNA fago T4 ($1,7 \times 10^5$ pb), e 1200 volte maggiore di quello del fago MS2. È chiaro quindi che la curva di Cot , o meglio il valore di $Cot_{1/2}$, ci danno importanti informazioni sulla complessità di un genoma.

Quando questo tipo di analisi viene applicato a DNA genomici di eucarioti superiori, la curva di riassociazione che si ottiene non è sigmoidale, ma presenta un andamento complesso con più gradini. Per esempio, la curva rossa in **figura 20.5** si riferisce al DNA genomico umano, che ha una lunghezza complessiva di circa $3,4 \times 10^9$ pb. L'andamento della curva di riassociazione indica chiaramente che, come abbiamo discusso nel capitolo 5, in questo genoma sono presenti diverse classi di DNA:

- circa il 40% del DNA riassocia con $Cot_{1/2}$ tra 10^3 e 10^4 ; esso viene chiamato **DNA unico**, o **non ripetitivo**, e contiene, tra l'altro, la quasi totalità dei geni veri e propri che sono presenti nel genoma in unica o pochissime copie;
- un altro 40-50% del DNA riassocia con $Cot_{1/2}$ tra 1 e 10, viene chiamato **DNA mediamente ripetitivo** (10^2 - 10^4) e comprende alcuni pochi geni che sono presenti in copie reiterate (ad es. i geni per l'rRNA e per i tRNA) oltre a trasposoni e retrotrasposoni attivi o non più attivi;
- infine, c'è una piccola frazione del genoma, 5-10%, costituita da sequenze semplici o **DNA satelliti**, costituiti da brevissime sequenze ripetute moltissime volte (10^5 - 10^6) nel genoma (*vedi* cap. 5).

Le tecniche di analisi della cinetica di riassociazione del DNA hanno subito numerosi cambiamenti e variazioni e hanno permesso di ottenere importanti informazioni sull'organizzazione dei genomi e, con alcuni adattamenti allo studio dell'RNA, anche sull'espressione genica. Non entreremo nel dettaglio di queste tecniche in quanto, come già detto, in buona parte superate dalla possibilità di analisi diretta delle sequenze nucleotidiche dei genomi e della loro espressione.

20.2 Uso di sonde di DNA per l'identificazione e analisi di sequenze nucleotidiche

La riassociazione di cui abbiamo parlato sopra si riferisce alla formazione in soluzione di molecole duplex tramite interazione dei due filamenti complementari precedentemente separati. Un uso un po' diverso della reazione di riassociazione di filamenti complementari viene fatto in tecniche in cui viene utilizzata una breve **sequenza sonda** (*probe*) per evidenziare e studiare la sua sequenza complementare in una complessa preparazione di DNA (o RNA). Sostanzialmente, la sonda, che è una sequenza nota e omogenea, viene usata come un reagente specifico che interagisce solo con la sequenza complementare all'interno di una popolazione eterogenea. La metodologia, che fu originalmente messa a punto e usata per studiare specifiche sequenze di DNA (geni) utilizzando un RNA come sonda, ha subito col tempo varie modifiche che oggi permettono di utilizzare sonde sia di DNA che di RNA per studiare bersagli che possono essere sia di DNA che di RNA. In tutti i casi, è necessario che la sonda sia in qualche modo visibile e/o misurabile. Ciò si ottiene marcando previamente la sonda con precursori radioattivi o colorati o fluorescenti (**finestra 20.1**).

Vi sono un paio di importanti differenze rispetto agli esperimenti di riassociazione del DNA prima descritti. Innanzitutto, in quel caso la concentrazione dei due filamenti che devono riassociarsi è per definizione uguale, in quanto prodotti per denaturazione di doppie eliche. Nel caso dell'uso di sonde, invece, la concentrazione può essere molto diversa. Spesso l'associazione viene fatta in presenza di un eccesso di sonda in modo da saturare la sequenza bersaglio che si vuole evidenziare e studiare. In alcuni casi, però, se quello che ci interessa è una

FINESTRA 20.1 MARCATURA DEL DNA CON ATOMI RADIOATTIVI

L'individuazione nelle librerie di DNA di cloni contenenti uno specifico gene ha tratto vantaggio dalla costruzione di specifiche sonde di DNA radioattive corrispondenti a segmenti del gene in esame. Tale sonda di DNA radioattiva, infatti, può ibridarsi, come descritto nel testo, con la parte del gene che ha la sequenza nucleotidica complementare.

Abbiamo già descritto, in figura 6.15, come l'accoppiamento tra l'attività polimerizzante 5'→3' della DNA polimerasi I di *Escherichia coli* con la sua attività esonucleasica 5'→3' permette la sintesi di molecole di DNA radioattive (**nick translation**) se nella miscela di reazione sono introdotti dNTP con il fosfato in posizione alfa marcato con l'isotopo [³²P].

Il metodo però più utilizzato per marcare radioattivamente un frammento di DNA in laboratorio è chiamato del **primer casuale** o **random priming**. Il metodo si basa, appunto, sull'ibridazione di una miscela di oligonucleotidi di tutte le possibili sequenze possibili (disponibile commercialmente) allo stampo denaturato del frammento di DNA

che si vuole marcare. La marcatura si ottiene aggiungendo uno o più dNTP radioattivi e un derivato della DNA polimerasi I di *E. coli* chiamato **frammento di Klenow**. Tale frammento è così chiamato perché Hans Klenow scoprì che è possibile tagliare la DNA polimerasi I in due frammenti: il più piccolo contiene solo il dominio proteico con l'attività esonucleasica 5'→3', mentre quello più grande (di Klenow) contiene sia l'attività polimerasica che l'attività esonucleasica di proofreading 3'→5' (vedi fig. 6.15).

È importante ricordare che per varie applicazioni (vedi, per esempio, la tecnica di DNA footprinting) è utile marcare il DNA alle estremità 5' con un atomo di fosforo radioattivo. Ciò si ottiene *in vitro* sfruttando le proprietà di un enzima chiamato **polinucleotide chinasi**. Tale enzima è in grado di catalizzare il trasferimento di un fosfato radioattivo in posizione γ dell'ATP (chiamato ATP γ [³²P]) e disponibile commercialmente) all'estremità 5' di molecole di DNA che hanno sia estremità 5'OH che estremità 5'-fosfato.

valutazione quantitativa della sonda, dovremo adottare condizioni di associazione in cui la sonda è limitante rispetto alla sequenza bersaglio.

Un'altra differenza è che, mentre nel caso della riassociazione di DNA denaturato i filamenti complementari si trovano ambedue in soluzione, nel caso dell'uso di sonde le sequenze bersaglio vengono spesso fissate su un supporto che viene poi fatto reagire con la sonda in soluzione. In molti casi, come per esempio il Southern blot e l'ibridazione *in situ*, questo è indispensabile perché ciò che ci interessa è proprio la distribuzione spaziale delle sequenze bersaglio. Ma anche in casi in cui ci interessa solo una titolazione quantitativa della sequenza bersaglio, si preferisce spesso fissare questa su un supporto che poi viene immerso nella soluzione contenente la sonda marcata. Il supporto maggiormente utilizzato per fissarvi le sequenze da analizzare è costituito da speciali filtri o membrane di nitrocellulosa o, più frequentemente, di nylon. Queste, che appaiono come dei fogli di carta, hanno la proprietà di legare il DNA denaturato (in certe condizioni anche l'RNA), permettendo però la sua interazione con la sonda e la formazione della doppia elica.

L'ibridazione o l'associazione di una sonda su una preparazione di DNA o RNA bersaglio sono alla base di molte tecniche diverse, alcune delle quali sono descritte appresso.

Ibridazione a saturazione

La saturazione del bersaglio con una sonda consente di misurare la percentuale del bersaglio complementare alla sonda stessa (**fig. 20.6**). Questa titolazione permette, per esempio, di "contare" il numero di copie di un gene, o di una sequenza non genica, all'interno di un genoma.

Ibridazione *in situ* o citologica

Questa tecnica permette di visualizzare dove si trova localizzata la sequenza bersaglio in strutture cellulari o istologiche. Le sonde possono essere radioattive, e quindi evidenziate mediante autoradiografia, oppure fluorescenti. È possibile visualizzare la localizzazione di specifiche sequenze genomiche in particolari re-

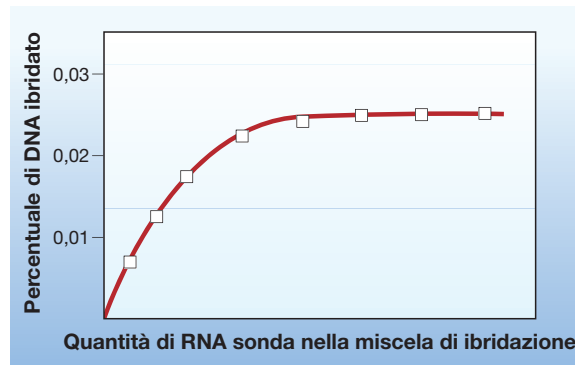
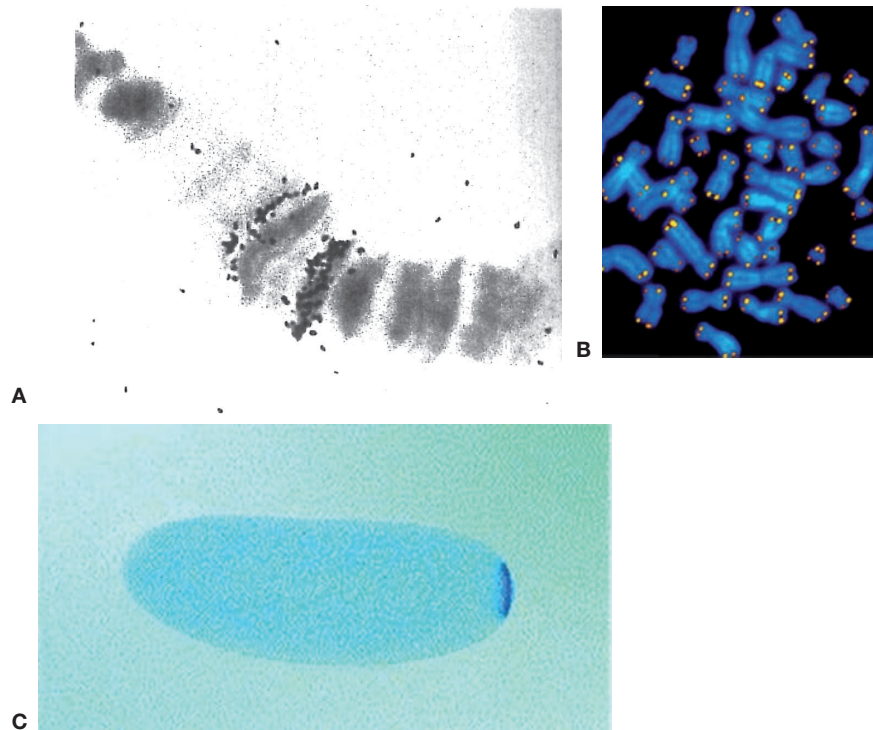


Figura 20.6 L'ibridazione a saturazione di una sonda su un DNA bersaglio permette di "contare" il numero di ripetizioni di una sequenza in un genoma. Esempio: determinazione del numero di geni per RNA ribosomale (rRNA) nel genoma umano. DNA genomico umano fissato su filtri di nitrocellulosa viene ibridato con una sonda radioattiva specifica per il gene che codifica per l'rRNA. Aumentando la concentrazione della sonda nella soluzione di ibridazione mantenendo fisso il tempo di reazione (oppure aumentando

il tempo di ibridazione mantenendo costante la concentrazione della sonda), la quantità di radioattività che si lega al DNA bersaglio aumenta fino a raggiungere il livello di saturazione quando tutte le sequenze ibridabili del DNA bersaglio sono effettivamente ibridate. Da questo valore, conoscendo la lunghezza del genoma e la lunghezza della sonda, è possibile risalire al numero di copie della sequenza nel DNA bersaglio e, quindi, al numero di geni per l'rRNA presenti in questo genoma.

Figura 20.7 Ibridazione *in situ* o "citologica". Esempi di ibridazione *in situ* per la localizzazione di sequenze di DNA o RNA in particolari strutture o distretti cellulari. (A) Una sonda radioattiva specifica per i geni attivati da shock termico è stata ibridata sul DNA denaturato di una preparazione di cromosomi politenici di *Drosophila* ed è stata successivamente evidenziata mediante autoradiografia dimostrando la localizzazione di questi geni su specifiche bande. (B) Una piastra metafase di cromosomi umani è stata analizzata con la tecnica chiamata FISH (Fluorescent In Situ Hybridization): il preparato è stato prima colorato con il DAPI (blu) che visualizza tutto il DNA cromosomico, poi è stato ibridato con una sonda fluorescente (giallo) specifica per la sequenza ripetuta TTAGGG dei telomeri. (C) La localizzazione al polo posteriore in un embrione di *Drosophila* dell'mRNA nanos è messa in evidenza mediante ibridazione *in situ* con una sonda specifica.



gioni del nucleo o nei cromosomi metafasici. Analogamente, si può indagare la localizzazione di specifici mRNA all'interno di una cellula o di un tessuto, come abbiamo discusso in altri capitoli del volume. Alcuni esempi di ibridazione *in situ* sono mostrati in **figura 20.7**.

Southern e Northern blot

Come sarà descritto più avanti, l'ibridazione con sonde specifiche permette l'analisi di molecole separate mediante elettroforesi su gel. Può trattarsi di frammenti

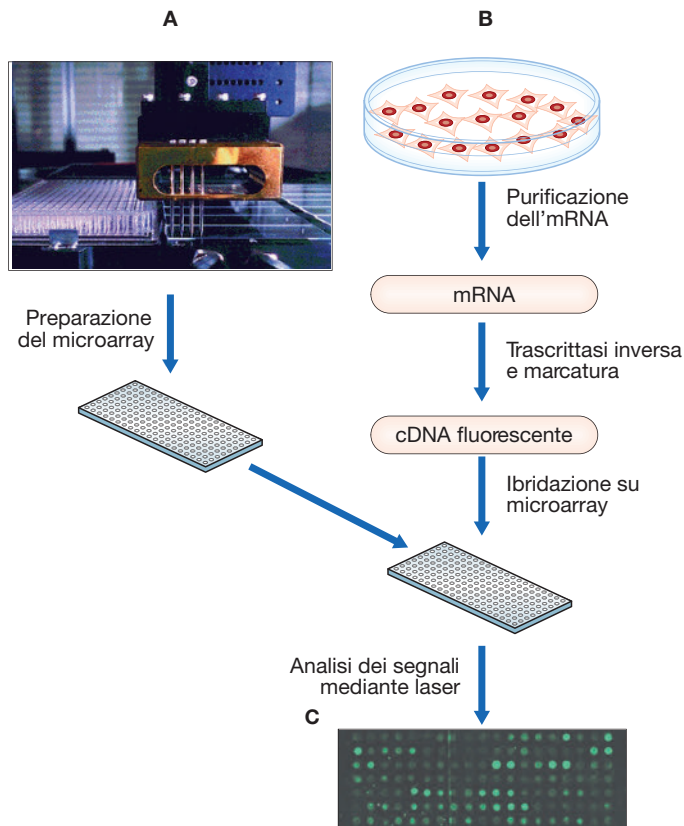


Figura 20.8 Tecnica dei microarray: schema della procedura. (A) Mediante un robot viene automaticamente preparato il microarray, cioè un supporto su cui sono allineati in maniera ordinata migliaia di campioni (sonde) di DNA rappresentativi dei diversi geni. (B) L'RNA estratto dalle cellule viene usato come stampo per la sintesi (ed eventualmente amplificazione mediante PCR) di cDNA fluorescente. (C) Il microarray viene immerso e incubato in una soluzione contenente il cDNA e successivamente lavato e analizzato per i segnali di fluorescenza.

di DNA ottenuti con enzimi di restrizione (Southern blot) o di uno specifico mRNA all'interno di una preparazione di mRNA totale (Northern blot). Una volta corsa l'elettroforesi, il DNA (o l'RNA) viene trasferito dal gel alla membrana e questa immersa in una soluzione contenente la sonda specifica per la sequenza bersaglio che vogliamo studiare.

Ibridazione su colonia o su placca

Come descritto più avanti, queste tecniche permettono di individuare le colonie batteriche, o le placche fagiche, in cui è presente una sequenza di interesse durante un esperimento di clonaggio.

Microarray (o microchip)

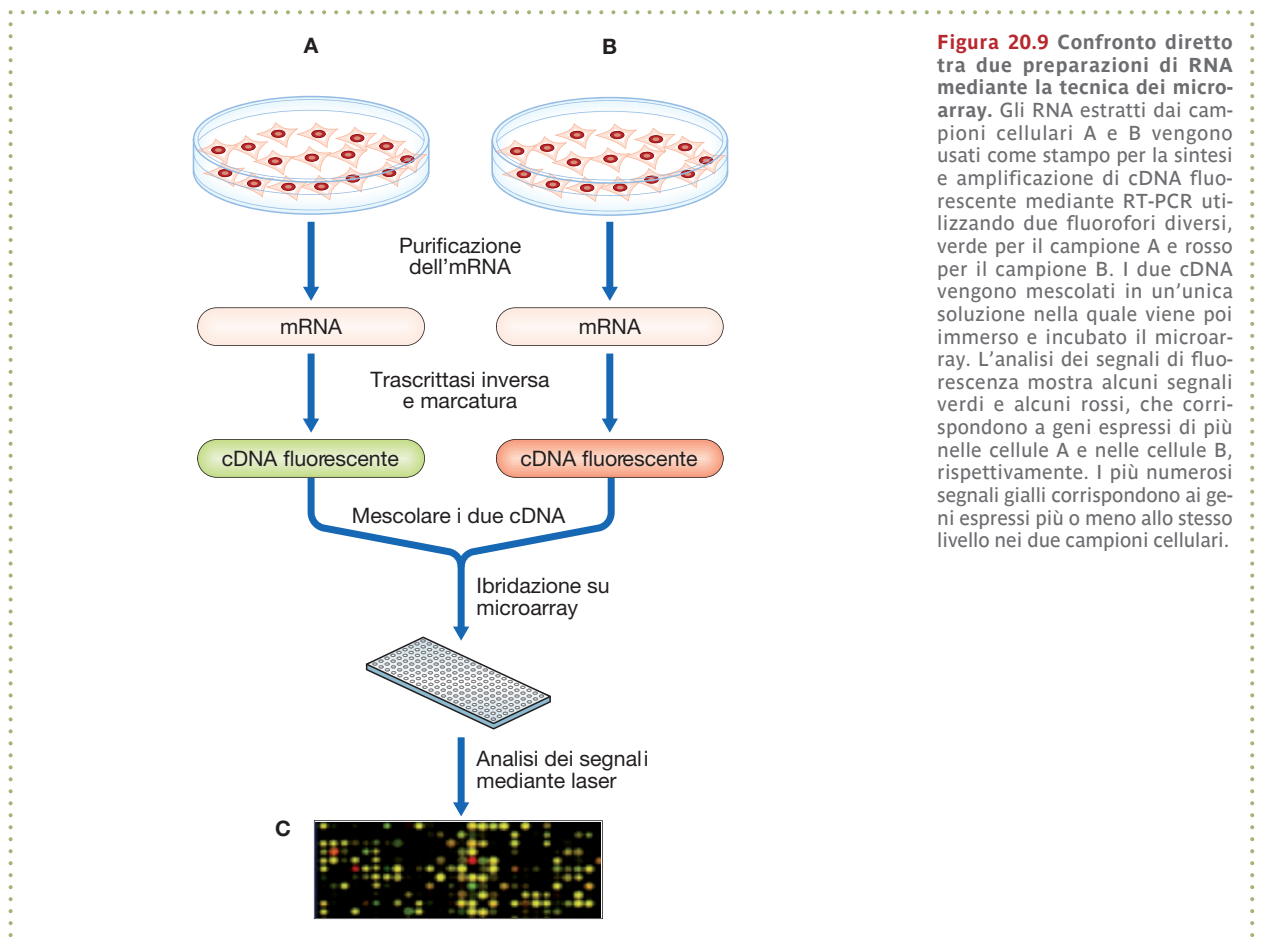
Ci soffermiamo un po' più a lungo su questa tecnica che rappresenta uno dei più recenti e interessanti sviluppi delle metodologie basate sull'associazione di sequenze nucleotidiche complementari. La tecnica dei microarray, schematizzata in **figura 20.8**, permette l'analisi quantitativa su vasta scala dell'espressione di numerosi geni, anche di un intero genoma (profilo di espressione). Un microarray consiste in un supporto, in genere di vetro o di silicio, su cui sono stati depositati (*spotted*) in maniera ordinata numerosi campioni di DNA di sequenza nota, corrispondenti ai diversi geni. Durante l'ibridazione (o associazione) questi legheranno specificamente le sequenze complementari che troveranno nella soluzione. Si noti che in questo caso l'uso di "sonda" e "bersaglio" è un po' diverso da quanto visto sopra: i campioni di DNA a sequenza nota fissati sul microarray sono sonde che servono a studiare le sequenze bersaglio

marcate (fluorescenti) che si trovano mescolate in soluzione. I campioni da depositare in maniera ordinata sui microarray, che rappresentano tutti i geni di cui ci interessa studiare l'espressione, possono essere ottenuti in vari modi: può trattarsi di cDNA clonati e purificati, o di frammenti genici ottenuti mediante amplificazione con PCR (*vedi* par. 20.6), o di oligonucleotidi sintetizzati chimicamente e poi depositati sul microarray oppure, con procedure più sofisticate, oligonucleotidi sintetizzati direttamente sul supporto. Le varie procedure sono state miniaturizzate e automatizzate, con la possibilità di depositare in maniera ordinata sul supporto migliaia di diversi campioni. Per la riassociazione, il microarray, preparato automaticamente da un robot, viene immerso in una soluzione contenente il materiale da analizzare. Questo può essere direttamente una preparazione di RNA, estratto dalle cellule o dal tessuto che ci interessa, marcato per fluorescenza. In genere però, per aumentare l'efficienza, la preparazione di RNA viene usata come stampo per una RT-PCR, che consiste nella sintesi di cDNA su stampo di RNA mediante trascrittasi inversa seguita da amplificazione mediante PCR, naturalmente in presenza di precursori fluorescenti per la marcatura delle molecole. La popolazione di cDNA così ottenuta è rappresentativa degli eventi trascrizionali nella cellula o tessuto di origine dell'RNA. A questo punto, e come mostrato in **figura 20.8**, si procede all'ibridazione immergendo e incubando i microarray nella soluzione di cDNA. Durante l'ibridazione i diversi cDNA fluorescenti riconoscono specificamente e legano le sequenze fissate in ordine sul microarray. Alla fine della reazione di associazione il microarray viene lavato, per allontanare il cDNA non ibridato, e poi analizzato per il segnale di fluorescenza. È importante notare che la preparazione bersaglio di cDNA (o mRNA) è un insieme molto eterogeneo di sequenze diverse, ciascuna delle quali è abbastanza diluita da non arrivare a saturare la sonda complementare fissata sul microarray; in queste condizioni le intensità dei segnali sui vari campioni depositati sul microarray ci danno una stima della concentrazione dei corrispondenti mRNA nella preparazione originale, che nel loro insieme costituiscono il **profilo d'espressione genica**.

In questo tipo di esperimenti i valori assoluti che si ottengono possono essere poco significativi e soggetti a vari errori. Pertanto, si preferisce in genere fare dei confronti di profili d'espressione genica in due condizioni diverse: per esempio, in cellule trattate con un inibitore e cellule di controllo non trattate, oppure in una linea cellulare tumorale e la corrispondente linea di controllo non tumorale, oppure in cellule prima e dopo il differenziamento ecc. In questi casi si applica la stessa procedura descritta sopra utilizzando, come mostrato in **figura 20.9**, due fluorofori diversi per i due tipi di cellule da confrontare, per esempio uno verde per il campione sperimentale A e uno rosso per il campione di controllo B. All'analisi dei risultati si osserveranno alcuni segnali verdi che indicheranno geni espressi maggiormente nel campione A, alcuni segnali rossi che indicheranno geni espressi maggiormente nel campione B, e segnali gialli che mettono in evidenza geni espressi più o meno allo stesso livello in ambedue i campioni (verde + rosso = giallo). In ogni caso, l'intensità del segnale ci dà una misura dell'espressione del gene corrispondente. Ovviamente, anche quest'ultima parte della procedura è stata automatizzata e computerizzata per permettere la misura del colore e dell'intensità di fluorescenza delle decine di migliaia di campioni dei microarray, l'immissione diretta dei dati in un computer e la loro analisi mediante speciali programmi appositamente elaborati.

20.3 Ultracentrifugazione

Centrifugazione e ultracentrifugazione sono tecniche di base nella pratica di laboratorio biologico. Esistono numerosi tipi di centrifughe e varie metodologie



per la purificazione e il frazionamento di diversi componenti cellulari e molecolari e per l'analisi delle loro proprietà e caratteristiche. Qui diamo per scontata una conoscenza di base della centrifugazione e tralasceremo molte tecniche. Ci limiteremo a presentarne due che sono state e sono molto utilizzate in Biologia Molecolare: la centrifugazione in “gradiente di saccarosio” e la centrifugazione in “gradiente di CsCl”. È importante distinguere bene questi due tipi di tecniche che, pur essendo superficialmente simili, sono molto diverse per l'informazione che danno.

Sedimentazione in gradienti di saccarosio

Tale tecnica permette di separare macromolecole e complessi macromolecolari in base alla loro velocità di sedimentazione.

La **centrifugazione in gradiente di saccarosio** permette di determinare con una certa approssimazione il coefficiente di sedimentazione (misurato in S = unità Svedberg) di una macromolecola, che dipende dalla massa (che è quello che di solito ci interessa misurare), ma anche dalla sua struttura e densità. Frequentemente la tecnica viene usata per analizzare il profilo di sedimentazione di macromolecole o di complessi macromolecolari, ed eventualmente frazionarli. In breve, il campione da analizzare viene depositato delicatamente sul gradiente di saccarosio preformato nella provetta e poi centrifugato per tempo e velocità opportuni (**fig. 20.10**). La centrifugazione viene interrotta quando molecole e complessi hanno migrato all'interno del gradiente che viene analizzato e/o raccolto in frazioni.

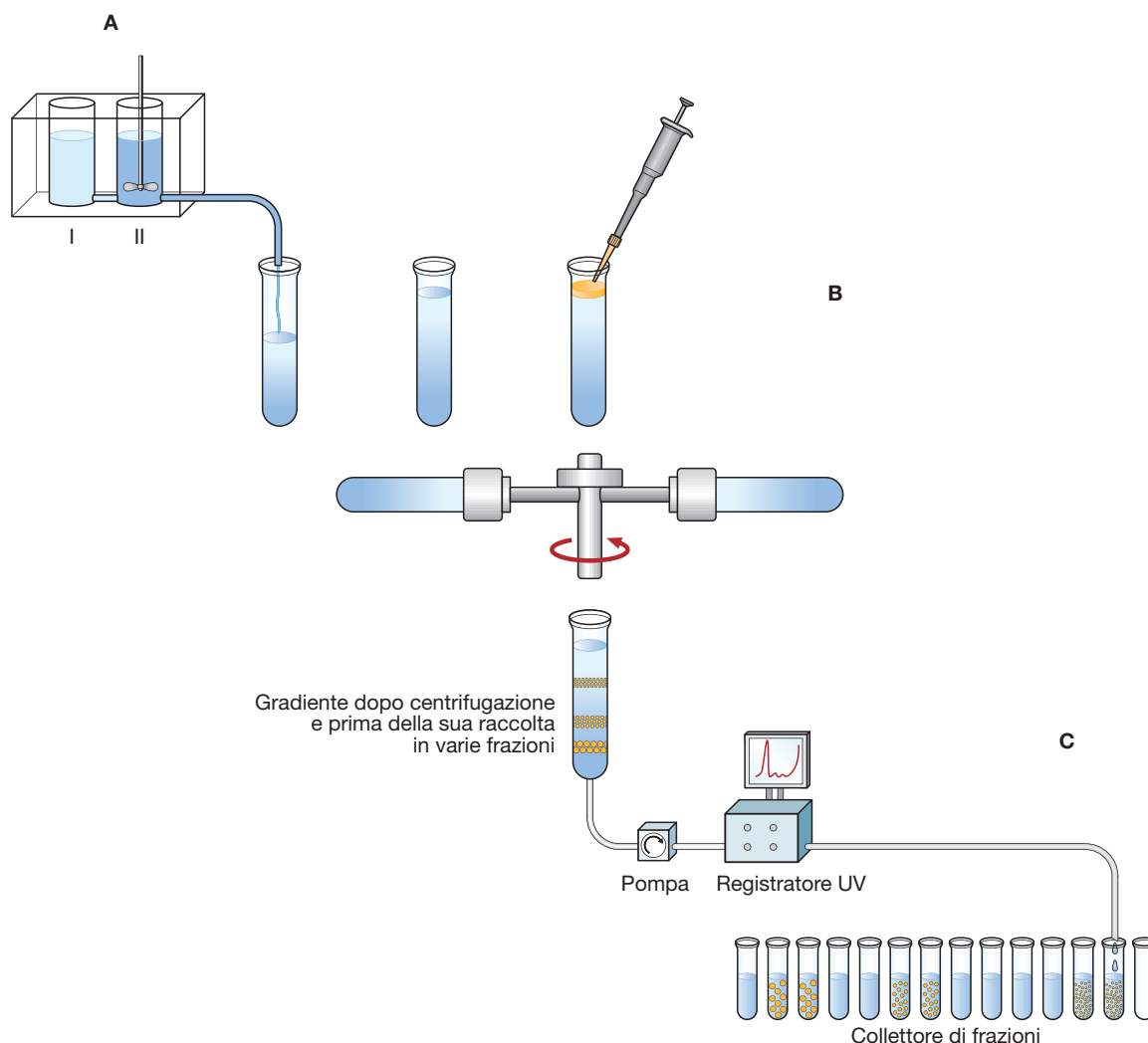


Figura 20.10 Sedimentazione in gradiente di concentrazione di saccarosio.

(A) *Preparazione di un gradiente di concentrazione di saccarosio.* Per la preparazione di un gradiente di saccarosio si utilizza un semplice apparato costituito da due vasi comunicanti, I e II. Nel comparto I viene messa la soluzione con concentrazione di saccarosio minore, per esempio 20%, e nel comparto II la soluzione più concentrata, per esempio 50%. Quando si inizia a fare defluire la soluzione dal comparto II alla provetta, viene aperta anche la comunicazione tra i due comparti. In questa maniera, mentre al fondo della provetta inizierà a stratificarsi il saccarosio concentrato (50%), contemporaneamente la soluzione diluita fluirà dal compartimento I al II, dove viene continuamente rimescolata da un agitatore. Il gradiente che si forma nella provetta è un gradiente lineare di concentrazione che va dal 50% al fondo al 20% in cima.

(B) *Sedimentazione in gradiente di concentrazione di saccarosio.* Il campione da analizzare viene caricato con una pipetta al di sopra di un gradiente di saccarosio, precedentemente preparato come mostrato in (A). La provetta contenente il gradiente viene centrifugata in un rotore di tipo swing-out, in cui l'asse della provetta viene a trovarsi sempre parallelo alla forza di gravità generata dalla centrifugazione. La centrifugazione viene interrotta quando il campione è sedimentato all'interno del gradiente, ma prima che sedimenti sul fondo.

(C) *Frazionamento di un gradiente di saccarosio.* Dopo la centrifugazione la provetta contenente il gradiente viene recuperata e si pratica un piccolo forellino alla base in modo che la soluzione inizi a gocciolare. La soluzione che gocciola viene prima fatta passare con una pompa attraverso uno spettrofotometro che ne registra l'assorbanza e poi frazionata in una serie di provette.

In questo tipo di esperimenti il ruolo principale del gradiente di saccarosio è quello di stabilizzare la colonna di liquido in cui si fanno sedimentare le molecole. Infatti, se non ci fosse il saccarosio, al momento della stratificazione del campione questo si rimescolerebbe con il liquido sottostante e, inoltre, diffusione e movimenti convettivi causerebbero un rapido rimescolamento del campione in tutta la provetta. Sfruttando il fatto che una soluzione a bassa concentrazione di saccarosio tende a restare stratificata sopra una soluzione a più alta concentrazione, la provetta per la sedimentazione viene preparata stratificando una serie

di soluzioni contenenti, dal basso verso l'alto, concentrazioni di saccarosio progressivamente decrescenti. Quindi, il gradiente di concentrazione di saccarosio ha lo scopo di stabilizzare la colonna di liquido, di permettere la stratificazione del campione al di sopra della colonna e di minimizzare movimenti convettivi e diffusione durante la centrifugazione. In realtà, per ottenere il gradiente non è necessario stratificare una alla volta le soluzioni a diversa concentrazione di saccarosio, ma si fa uso di un semplice strumento descritto in **figura 20.10A**.

Come già detto, dopo avere preparato il gradiente e averci caricato sopra il campione, la provetta viene centrifugata ad alta velocità. Questa viene scelta a seconda del materiale che analizziamo e dello scopo che ci prefiggiamo; comunque si tratta di velocità tra circa 20 000 e 40 000 rpm (*rounds per minute* = giri al minuto), corrispondenti, a seconda del raggio del **rotore**, a circa 30 000-100 000 g (gravità). Per questi esperimenti è necessario usare un rotore da centrifuga di tipo particolare, chiamato *swing-out*, in cui i supporti contenenti le provette sono appesi con delle cerniere al rotore in modo che, inizialmente verticali, durante l'accelerazione assumano una posizione orizzontale (**fig. 20.10B**). In questo modo, l'asse della provetta resta sempre orientato nella direzione della forza centrifuga. Durante la centrifugazione le macromolecole e i complessi migrano attraverso la colonna di liquido e andrebbero a sedimentare sul fondo; invece, al momento opportuno, quando si presume che il materiale di interesse sia migrato sufficientemente ma prima che abbia raggiunto il fondo, la centrifugazione, che può durare da decine di minuti a qualche ora, viene interrotta. Quando il rotore si ferma, i supporti contenenti le provette ritornano lentamente in posizione verticale e le provette vengono recuperate per l'analisi del risultato. Il modo più semplice è di praticare con un ago un piccolo foro sul fondo della provetta (di plastica) da cui il gradiente inizierà a gocciolare: prima gli strati di saccarosio più concentrato che si trovano al fondo della provetta e poi, progressivamente, gli strati superiori. Mano a mano che la soluzione esce, viene raccolta in un certo numero di frazioni (provette). In questo modo, le macromolecole o complessi che erano sedimentati ad altezze diverse del gradiente andranno a trovarsi in diverse frazioni. In alcuni casi si può preferire seguire con uno spettrofotometro l'andamento del profilo di sedimentazione (**fig. 20.10C**). In questo caso anziché raccoglierlo direttamente nelle provette, il gradiente viene fatto passare con un sottile tubo attraverso la cuvetta di uno spettrofotometro che legge in modo continuo l'assorbanza a 260 nm, producendo un grafico che descrive il profilo di sedimentazione; la soluzione che esce dallo spettrofotometro viene raccolta in frazioni come detto sopra. La **figura 20.11** mostra un esempio di profilo di sedimentazione ottenuto mediante centrifugazione in gradiente di saccarosio di

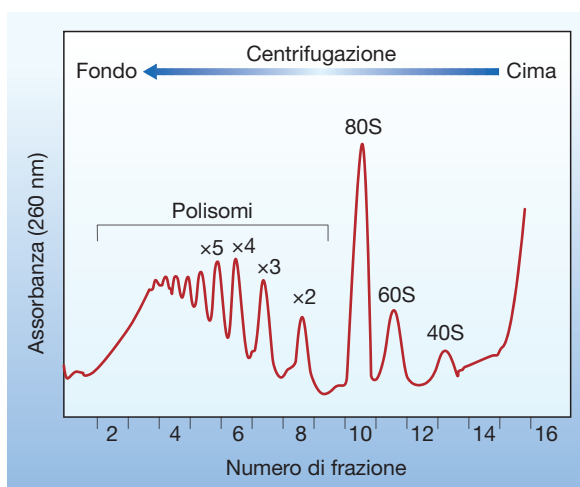


Figura 20.11 Esempio di analisi del profilo di sedimentazione in gradiente di saccarosio. Un estratto citoplasmatico di cellule eucariotiche è stato centrifugato in un gradiente di saccarosio 25-70% per 2 ore a 37 000 rpm. Alla fine della centrifugazione il gradiente è stato fatto defluire dal fondo e, prima della raccolta in frazioni, è stato fatto passare attraverso la cuvetta di uno spettrofotometro che ne ha registrato il profilo di assorbanza (a 260 nm). È possibile osservare la sedimentazione differenziale dei ribosomi 80S e delle due subunità 60S e 40S, e dei polisomi costituiti da 2, 3, 4 ecc. ribosomi.

un estratto citoplasmatico di cellule di mammifero in cui si separano i polisomi dai singoli ribosomi e dalle loro subunità.

Gradienti di densità di cloruro di cesio (CsCl)

Brevemente, mediante tale tecnica si analizzano e frazionano macromolecole e complessi macromolecolari sulla base della loro densità.

Questo tipo di **centrifugazione** viene anche chiamata **isopicnica** in quanto il campione non sedimenta dalla cima al fondo della provetta ma va a formare una banda nella posizione del gradiente che ha la sua stessa densità, che, per questo, è chiamata **densità di galleggiamento** (in inglese *buoyant density*). Sia la procedura che il principio su cui si basa sono molto diversi da quanto descritto sopra per i gradienti di saccarosio. Mentre la sedimentazione in gradiente di saccarosio discrimina macromolecole e complessi sulla base del coefficiente di sedimentazione, che in buona parte dipende dalla massa, la centrifugazione in gradiente di CsCl discrimina diversi DNA (ma anche RNA) sulla base della densità (peso specifico), indipendentemente dal loro peso molecolare. Inoltre, a differenza dei gradienti di saccarosio che venivano preformati prima del caricamento del campione, nel caso dei gradienti di CsCl si inizia, come illustrato in **figura 20.12**, con una provetta

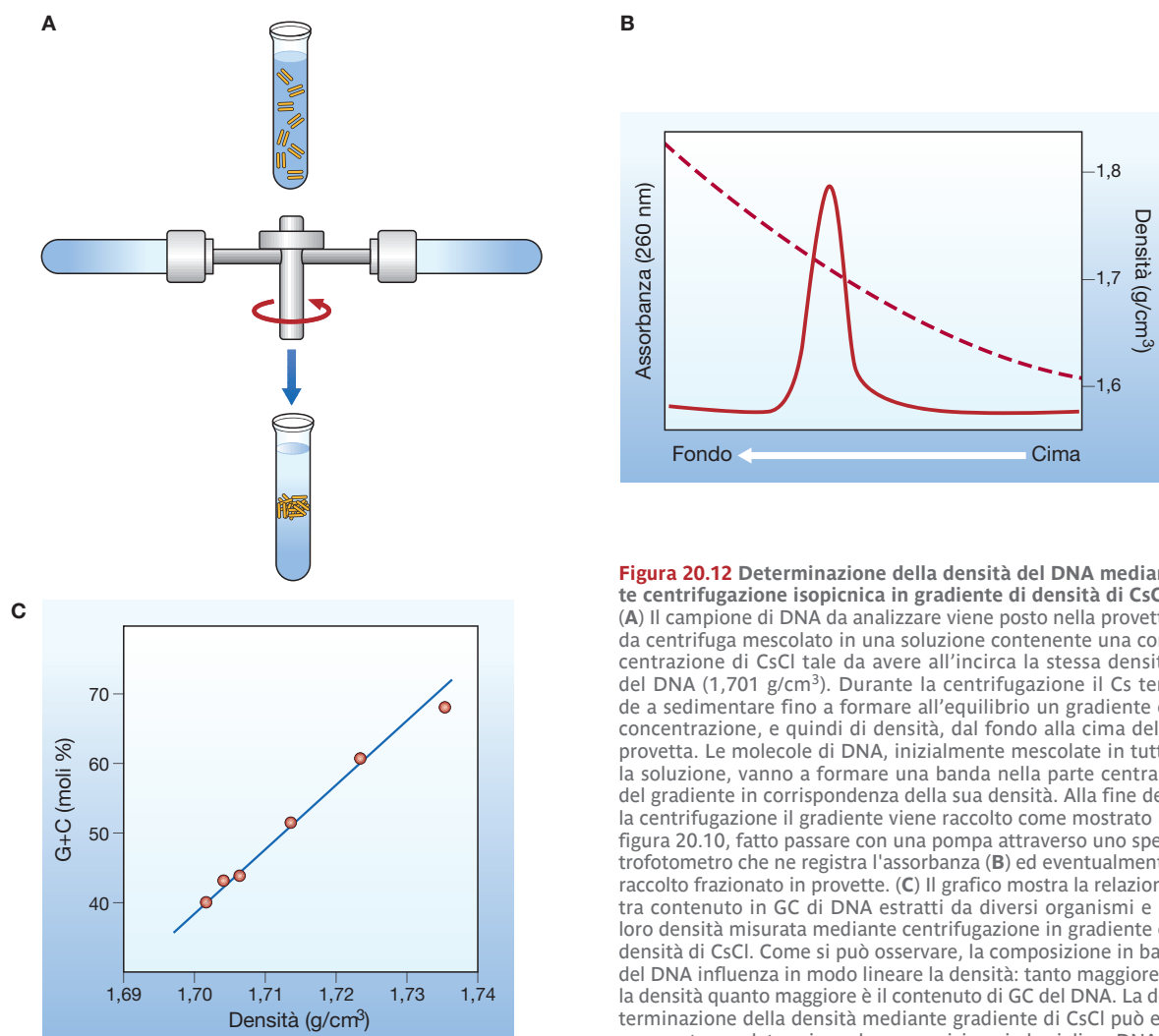


Figura 20.12 Determinazione della densità del DNA mediante centrifugazione isopicnica in gradiente di densità di CsCl.

(A) Il campione di DNA da analizzare viene posto nella provetta da centrifuga mescolato in una soluzione contenente una concentrazione di CsCl tale da avere all'incirca la stessa densità del DNA (1,701 g/cm³). Durante la centrifugazione il Cs tende a sedimentare fino a formare all'equilibrio un gradiente di concentrazione, e quindi di densità, dal fondo alla cima della provetta. Le molecole di DNA, inizialmente mescolate in tutta la soluzione, vanno a formare una banda nella parte centrale del gradiente in corrispondenza della sua densità. Alla fine della centrifugazione il gradiente viene raccolto come mostrato in figura 20.10, fatto passare con una pompa attraverso uno spettrofotometro che ne registra l'assorbanza (B) ed eventualmente raccolto frazionato in provette. (C) Il grafico mostra la relazione tra contenuto in GC di DNA estratti da diversi organismi e la loro densità misurata mediante centrifugazione in gradienti di densità di CsCl. Come si può osservare, la composizione in basi del DNA influenza in modo lineare la densità: tanto maggiore è la densità quanto maggiore è il contenuto di GC del DNA. La determinazione della densità mediante gradiente di CsCl può essere usata per determinare la composizione in basi di un DNA.

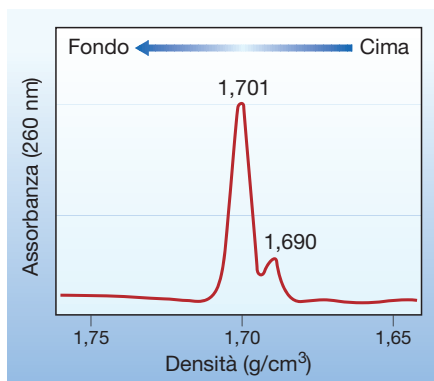


Figura 20.13 Analisi del DNA di topo mediante centrifugazione in gradiente di CsCl. Il DNA di topo sottoposto a centrifugazione in un gradiente di densità di CsCl si separa in due bande: una banda principale con densità di circa 1,701 g/cm³ che corrisponde alla maggior parte del genoma di topo, e una banda minore, più leggera, con densità di circa 1,690 g/cm³ che costituisce il DNA satellite.

in cui si trovano mescolati sia il DNA che il CsCl. L'atomo di Cs è molto pesante, avendo peso atomico di circa 133 (da confrontare con circa 23 del sodio), e, sotto l'effetto delle 100 000 g create dall'ultracentrifugazione, tende a sedimentare. L'equilibrio che si crea tra la tendenza a sedimentare e la diffusione, porta alla formazione di un gradiente di concentrazione del Cs⁺ tra il fondo e la cima della provetta da centrifuga, a cui corrisponde un gradiente di densità. La concentrazione del CsCl nella soluzione di partenza viene scelta in modo che questa abbia una densità vicina a quella del DNA (circa 1,701 g/cm³). Dopo che per effetto della centrifugazione si è formato il gradiente, la soluzione avrà la densità di circa 1,701 a metà del gradiente, mentre avrà una densità maggiore al fondo e una minore alla cima della provetta. Nel frattempo, mano a mano che durante la centrifugazione si forma il gradiente di Cs⁺, il DNA che inizialmente si trovava mescolato nella soluzione si sposta nella provetta, andando "a fondo" rispetto al Cs⁺ meno denso e "a galla" rispetto a quello più denso, fino a formare una banda all'altezza del gradiente dove trova la sua stessa densità. La formazione del gradiente di CsCl e della banda di DNA rappresentano un equilibrio che viene raggiunto dopo tempi di centrifugazione piuttosto lunghi, in genere 48 ore.

Dalla densità del cesio in cui si forma la banda si ricava la densità del DNA, e da questa si può calcolare la composizione in basi del DNA. Infatti, come mostrato in **figura 20.12C**, esiste una relazione lineare tra densità di un DNA e il suo contenuto in GC, tanto che la misura della densità ha rappresentato per molti anni uno dei due metodi utilizzati per determinare la composizione in basi del DNA (l'altro metodo è la misura della T_m di cui abbiamo parlato precedentemente).

La **figura 20.13** mostra come l'uso della centrifugazione in gradiente di CsCl permette anche di individuare e frazionare "DNA satelliti" nei genomi eucariotici.

La centrifugazione in gradienti di densità di CsCl può essere utilizzata anche in esperimenti in cui la macromolecola di interesse viene marcata con isotopi pesanti degli atomi. Un importante esperimento storico di questo tipo è quello con cui Meselson e Stahl dimostrarono la replicazione semiconservativa del DNA, utilizzando l'incorporazione dell'isotopo pesante (¹⁵N) e dell'isotopo normale (¹⁴N) dell'azoto per marcare, rispettivamente, i filamenti parentali e quelli di nuova sintesi del DNA, come abbiamo descritto nel capitolo 6 (*vedi par. 6.1*). In realtà per quell'esperimento è stata usata una centrifuga analitica (**finestra 20.2**), che, anche se non permetteva il frazionamento delle varie bande, ne permetteva un'accurata misurazione della densità.

Gradienti di densità di CsCl vengono usati anche per separare DNA circolari covalentemente chiusi da DNA lineari: per esempio DNA plasmidico rispetto a genomico, o DNA mitocondriale rispetto a nucleare. Infatti, le molecole di DNA circolari superavvolte hanno una densità maggiore rispetto alle molecole lineari. Inoltre, la differenza di densità può essere ulteriormente aumentata dall'aggiunta

FINESTRA 20.2

ULTRACENTRIFUGA ANALITICA

La centrifuga di cui abbiamo parlato nel testo, e comunemente usata nella maggior parte dei laboratori, è l'ultracentrifuga "preparativa". In realtà, la velocità di sedimentazione e la densità di molecole e complessi possono essere studiate più accuratamente mediante l'ultracentrifuga "analitica", che combina una normale ultracentrifuga con un apparato spettrofotometrico con cui si può osservare in "tempo reale" cosa succede all'interno della cella durante la centrifugazione. Infatti, la cella di centrifugazione attraversa il rotore ed è chiusa, sopra e sotto, da pareti di quarzo attraverso le quali il sistema ottico può rilevare il materiale che sedimenta o che forma una banda, basandosi sull'assorbimento della

luce ultravioletta, oppure registrando variazioni dell'indice di rifrazione. A intervalli regolari di tempo, durante la centrifugazione l'operatore può scattare fotografie o acquisire immagini con un computer, seguendo così l'evoluzione del campione durante la sedimentazione verso il fondo o durante la formazione di bande isopicniche. In questa maniera la misura della velocità di sedimentazione o della posizione di una banda nel gradiente è molto più accurata che in una centrifuga preparativa. Lo svantaggio di questo strumento, oltre alla complessità e ai costi, sta nel fatto che alla fine dell'analisi, quando si ferma la centrifuga, il materiale si rimischia e non può essere raccolto o frazionato.

Tabella 20.1 Confronto schematico tra le due tecniche di ultracentrifugazione: sedimentazione in gradiente di saccarosio e galleggiamento in gradiente di densità di CsCl.

	Gradiente di saccarosio	Gradiente di CsCl
Proprietà rilevante	Velocità di sedimentazione	Densità
Scopo del gradiente	Stabilizzare colonna liquido	Formare un gradiente di densità
Preparazione del gradiente	Preformato	Si forma durante la centrifugazione
Inserimento del campione	Depositato in cima al gradiente	Inizialmente mescolato in tutta la soluzione
Tempo di centrifugazione	Deve essere limitato, altrimenti il materiale sedimenta sul fondo	C'è un tempo minimo necessario per la formazione del gradiente e delle bande. Il materiale non sedimenta mai sul fondo

di bromuro di etidio alla soluzione. Questo si intercala nella doppia elica del DNA srotolandola un po' e quindi rendendola meno densa. Poiché in un DNA circolare covalentemente chiuso si può intercalare meno bromuro d'etidio che in un DNA lineare, quest'ultimo risulterà più leggero di un DNA chiuso in un gradiente contenente bromuro d'etidio.

La **tabella 20.1** riassume e mette a confronto le caratteristiche salienti delle due tecniche, sedimentazione in gradienti di saccarosio e gradienti di densità di CsCl.

20.4 Tecnologie di base per l'isolamento e la manipolazione di geni

Abbiamo discusso nel testo la complessità dei meccanismi che controllano l'espressione dei geni determinando il fenotipo di una cellula e il suo differenziamento. La comprensione a livello molecolare di fenomeni biologici così complessi ha richiesto l'utilizzo di approcci sperimentali molto efficienti e tra loro integrati e complementari.

Certamente lo sviluppo più sconvolgente nell'identificazione dei geni e dei meccanismi che ne controllano l'espressione lo si è avuto negli ultimi 30-40 anni con la messa a punto di tecnologie capaci di sezionare il genoma, identificare i geni, analizzarli e modificarli. L'insieme di tali tecnologie ha preso il nome di **Ingegneria Genetica** o **Tecnologie del DNA ricombinante**. La capacità di isolare i

geni e di capire a livello molecolare come la loro espressione possa essere regolata ha completamente rivoluzionato i campi della Genetica e della Biochimica e, di fatto, coincide con lo sviluppo della Biologia Molecolare. Va al di là dello scopo di questo testo fornire una descrizione dettagliata delle numerose tecnologie applicate allo studio della struttura e funzione dei geni: al riguardo, rimandiamo ad alcuni testi specializzati. In questo paragrafo cercheremo di descrivere come le tecnologie del DNA ricombinante si siano via via sviluppate con l'accrescersi delle nostre conoscenze teoriche e limiteremo la nostra presentazione alle tecniche e applicazioni più rilevanti e connesse con gli argomenti trattati nel testo.

La rivoluzione determinata dallo sviluppo delle tecnologie del DNA ricombinante non è legata allo sviluppo di una singola scoperta teorica, ma piuttosto alla combinazione di una varietà di conoscenze e relative applicazioni. Il prototipo ormai classico di un esperimento di ingegneria genetica è connesso al **clonaggio di un segmento di DNA**. Riferito a frammenti di DNA, il termine "clonaggio" indica l'introduzione di un gene (o di un suo segmento) all'interno di una cellula, in modo che tale frammento venga copiato quando la cellula si replica, determinando la produzione di numerose cellule tutte contenenti almeno una copia di quel frammento, che si potrà così dire essere stato clonato. In una tale accezione, un esperimento di clonaggio di DNA è diviso in due parti: una parte *in vitro* in cui si producono frammenti di DNA per inserirli in un opportuno elemento genetico (vettore di clonaggio) generando, così, una molecola di DNA ricombinante (**fig. 20.14**); una seconda parte *in vivo* che consiste nell'introduzione della mole-

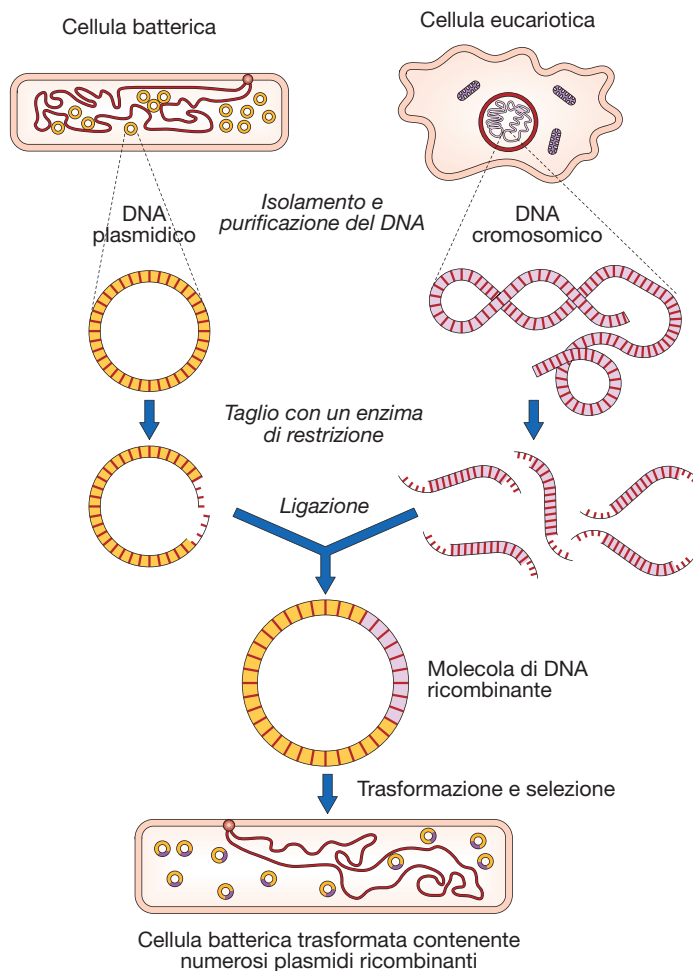


Figura 20.14 Schema di un tipico esperimento di clonaggio. Lo schema presentato si riferisce al clonaggio di frammenti di DNA di un organismo eucariotico all'interno di un vettore di clonaggio plasmidico e successiva trasformazione e selezione delle cellule di *Escherichia coli* trasformate contenenti molecole di DNA ricombinanti.

FINESTRA 20.3

UTILIZZO DELL'ELETTROFORESI PER SEPARARE MOLECOLE DI ACIDI NUCLEICI E PROTEINE

Molecole diverse di DNA e proteine possono essere separate mediante elettroforesi su gel perché differiscono nella loro carica elettrica, dimensione e struttura. Il gel è una matrice solida, per lo più costituita da agarosio o acrilammide:

- l'**agarosio** è uno zucchero solubile in acqua alla temperatura di ebollizione che diventa solido quando si raffredda, formando una matrice attraverso dei legami idrogeno tra le catene lineari;

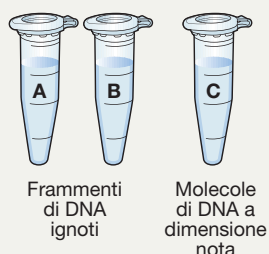
- i gel di **poliacrilammide** si formano, invece, tramite copolimerizzazione di acrilammide, un monomero solubile in acqua, e di un agente che forma legami trasversali così da formare un reticolo tridimensionale.

I due tipi di gel differiscono per le dimensioni dei pori del reticolo: gel con pori grandi permettono la separazione di molecole di grandi dimensioni, gel con pori piccoli separano molecole di dimensioni limitate e, come si vedrà nelle rea-

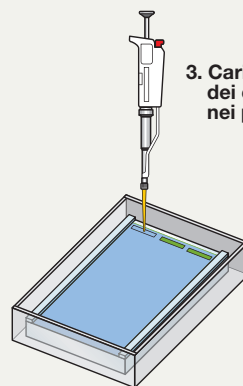
1. Preparazione del gel



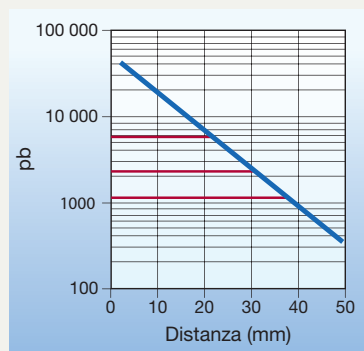
2. Campioni di DNA da analizzare



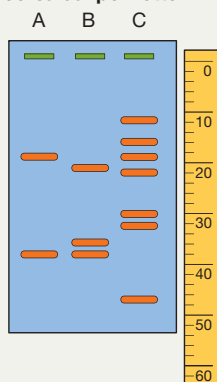
3. Caricamento dei campioni di DNA nei pozzetti del gel



6. Determinazione della curva standard per determinare le dimensioni di frammenti di DNA



5. Colorazione delle molecole di DNA con bromuro di etidio e misurazione della distanza percorsa dal pozzetto



4. Separazione dei frammenti di DNA tramite elettroforesi

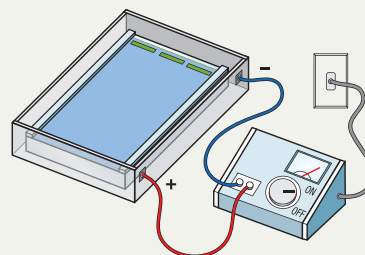
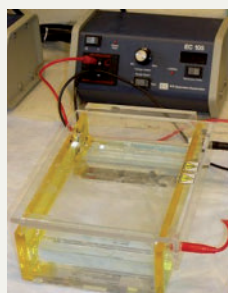
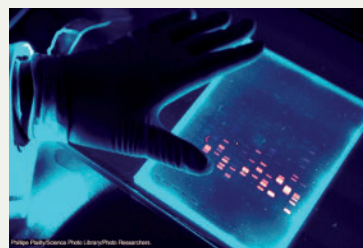


Figura 20F3.1 Separazione di molecole di DNA di dimensioni diverse mediante elettroforesi su gel di agarosio. (A) Procedura schematica per la separazione elettroforetica di molecole di DNA di diverse dimensioni su un gel di agarosio orizzontale. La posizione dei frammenti di DNA di diverse dimensioni è visualizzata dopo colorazione con bromuro d'etidio e la distanza di migrazione percorsa da molecole a dimensione nota (corsia di destra in 5.C) permette la costruzione di una curva standard per determinare le dimensioni di molecole di DNA di dimensione ignota (corsie di sinistra e centrale). (B) Nella parte in basso a sinistra della figura è mostrato un tipico alimentatore per la creazione del campo elettrico richiesto per l'elettroforesi e una tipica cella elettroforetica per la separazione su gel di agarosio orizzontale. Si notino i cavi elettrici che definiscono la posizione del polo negativo o catodo (nero) e del polo positivo o anodo (rosso). (C) Molecole



B C



di DNA separate mediante elettroforesi su gel di agarosio sono visualizzate dopo colorazione con bromuro d'etidio e illuminazione con una lampada a raggi UV.

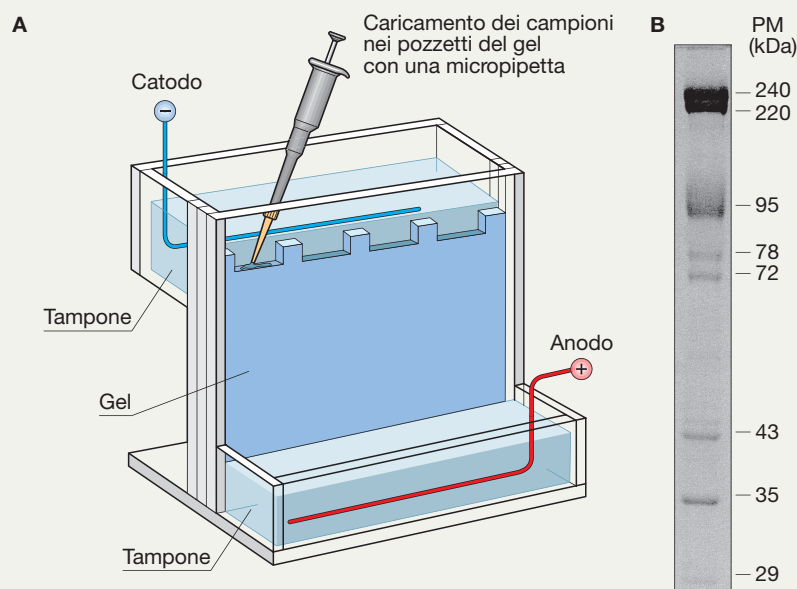


Figura 20F3.2 Separazione di proteine di dimensioni diverse mediante elettroforesi SDS-PAGE. (A) Disegno schematico di un apparecchio verticale per la separazione di proteine. (B) Separazione di proteine di diverso Peso Molecolare (PM) indicato in kiloDalton (kDa) mediante SDS-PAGE e colorazione con Coomassie blue.

zioni di sequenziamento del DNA, molecole di DNA che differiscono una dall'altra per un singolo nucleotide.

Le tecniche più utilizzate per la separazione di molecole di acidi nucleici e proteine variano sensibilmente e sono qui descritte separatamente.

Se un gel è sottoposto all'azione di un campo elettrico, molecole di DNA e RNA migrano nel gel in funzione della loro carica, dimensione e struttura. La **figura 20F3.1B** mostra un tipico apparecchio per la separazione orizzontale di molecole di DNA o RNA in un gel di agarosio. Come si può notare, l'apparecchio per l'elettroforesi è collegato a un alimentatore di corrente elettrica. Le molecole di DNA e RNA sono cariche negativamente per la presenza di gruppi fosfato e migrano, quindi, dal polo negativo (catodo) verso il polo positivo (anodo). La **figura 20F3.1A** dà una rappresentazione visiva dei diversi passaggi richiesti per la separazione di molecole di DNA di diverse dimensioni in un gel di agarosio. Frammenti di DNA lineari di diverse dimensioni generati, per esempio, con il taglio da parte di enzimi di restrizione sono separabili tramite elettroforesi su gel di agarosio. Come mostrato in figura, disponendo di frammenti di DNA lineari a lunghezza nota, è possibile costruire su una carta semilogaritmica una curva standard dove in ascissa è riportata la distanza di migrazione dei frammenti noti e in ordinata la loro dimensione in paia di basi. La retta così generata permette di estrapolare la dimensione di frammenti di dimensione ignota dopo averne misurato la distanza di migrazione nel gel. I frammenti di DNA che si separano tramite elettroforesi possono essere facilmente visualizzati immergendo il gel in una soluzione contenente bromuro d'etidio. Tale sostanza si intercala tra i due filamenti di DNA ed emette una fluorescenza di colore arancione (**fig. 20F3.1C**) quando esposta a raggi UV, permettendo di visualizzare i frammenti di DNA nel gel.

Molecole di DNA si separano per elettroforesi anche in base alla loro struttura: per esempio, il DNA circolare di un plasmide migra diversamente da una molecola di DNA lineare della stessa dimensione.

Come descritto nel paragrafo 2.4, una molecola di DNA circolare può essere rilassata o superavvolta. Una molecola di DNA circolare rilassata ha una struttura che determina il maggior rallentamento nella migrazione elettroforetica e, quindi, migra più lentamente della stessa molecola di DNA linearizzata, per esempio, mediante un singolo taglio con enzima di restrizione. Al contrario, la stessa molecola di DNA circolare superavvolta, essendo molto più compatta, migra più velocemente della molecola di DNA con struttura lineare o circolare rilassata.

L'elettroforesi su gel permette anche di separare molecole di RNA di diverse dimensioni in modo del tutto analogo a quanto descritto sopra per molecole di DNA.

L'elettroforesi in un campo elettrico è la tecnica più utilizzata anche per separare tra loro miscele di proteine di diverse dimensioni. La separazione dei vari componenti proteici avviene solitamente su gel verticali di poliacrilammide (**fig. 20F3.2A**), con la porosità più confacente alle dimensioni delle proteine da separare.

La velocità di migrazione su gel di una proteina dipende dall'intensità del campo elettrico applicato, dalla dimensione e dalla carica. La carica netta di una proteina dipende, a sua volta, dal pH del tampone utilizzato per la corsa elettroforetica. Non rientra nello scopo di questo manuale offrire una descrizione dettagliata delle numerose tecniche elettroforetiche per la separazione di proteine. Ci limitiamo a considerare che tali tecniche possono essere divise in due categorie principali: elettroforesi di proteine in condizioni native o in condizioni denaturanti. **In condizioni native** le proteine conservano la loro conformazione e la separazione avviene sulla base del rapporto carica/massa. La tecnica più utilizzata in laboratorio è, però, la separazione di proteine **in condizioni denaturanti** in presenza di sodio dodecil solfato (SDS): si parla di tecnica

(continua)

SDS-PAGE (Sodio Dodecil Solfato-PoliAcrilammide Gel Elettroforesi). In presenza di SDS e di altre condizioni sperimentali specifiche, le strutture secondarie delle proteine sono denaturate e le molecole vengono di fatto ricoperte da cariche negative permettendo la migrazione delle proteine verso l'anodo con una velocità essenzialmente dipendente dalle dimensioni (peso molecolare) delle diverse specie proteiche presenti nella miscela. Ne consegue che,

in modo analogo a quanto descritto sopra per il DNA, disponendo di proteine di dimensione nota è possibile costruire una curva standard che permette di calcolare il peso molecolare di una proteina ignota in funzione della sua distanza di migrazione dopo elettroforesi. Anche le proteine possono essere visualizzate dopo la corsa elettroforetica utilizzando vari coloranti, il più comune dei quali è il Coomassie blue (**fig. 20F3.2B**).

cola di DNA ricombinante all'interno di un organismo in cui questa sia in grado di replicarsi autonomamente o di perpetuarsi integrandosi all'interno del genoma.

Le molecole di DNA sono generalmente molto grandi ed è quindi impossibile clonarle come tali all'interno di un vettore di clonaggio, i cui tipi più comuni saranno descritti appresso. Se, quindi, vogliamo clonare singoli geni o frammenti di geni dal genoma di uno specifico organismo, dobbiamo estrarre il DNA dall'organismo in esame e procedere alla sua frammentazione in molecole che abbiano dimensioni gestibili sperimentalmente. Il DNA può essere rotto in pezzi casuali di dimensioni più piccole attraverso una serie di manipolazioni meccaniche, la più frequente delle quali è l'utilizzo di ultrasuoni: il DNA, infatti, viene frammentato in segmenti sempre più piccoli aumentando la durata o l'intensità del trattamento. Come abbiamo descritto precedentemente, le dimensioni di molecole di DNA possono essere determinate con varie tecnologie; la più utilizzata è oggi l'**elettroforesi su gel** (**finestra 20.3**).

Il genoma di un organismo può, però, essere frammentato non a caso, ma in corrispondenza di specifiche sequenze sul DNA. La scoperta di enzimi in grado di tagliare il DNA in corrispondenza di particolari sequenze nucleotidiche è stata, forse, la scoperta singolarmente più importante per lo sviluppo di tutte le tecnologie del DNA ricombinante: tali proteine sono chiamate **enzimi di restrizione**.

Enzimi di restrizione e DNA ligasi

Biochimicamente, gli enzimi di restrizione (**ER**) sono delle endonucleasi che provocano delle rotture interne a doppio filamento sul DNA in corrispondenza di particolari sequenze nucleotidiche.

L'utilizzo degli ER è ormai entrato nella pratica corrente delle tecnologie del DNA ricombinante, ma la loro scoperta è avvenuta tramite lo studio molecolare di un processo biologico che era molto lontano dall'idea di una sua potenziale applicazione pratica. I ricercatori avevano appreso che un certo batteriofago determina la lisi, cioè la morte, di cellule batteriche infettate dal virus con percentuali molto diverse da un ceppo all'altro di batteri: la capacità di un ceppo batterico di prevenire la replicazione di un virus infettante era stata chiamata **restrizione dell'ospite**. La restrizione, però, non è assoluta: può accadere che un certo virus causi la lisi di un numero limitatissimo di cellule ospiti; tuttavia, se questi batteriofagi vengono recuperati e utilizzati per re-infettare lo stesso ceppo di batteri utilizzato precedentemente, si trova che l'efficienza d'infezione è adesso diventata altissima e si parla di modificazione dell'ospite. Quali meccanismi sono alla base del sistema di modificazione e restrizione dell'ospite? Si è osservato che, dopo l'infezione, il DNA del batteriofago è tagliato in piccoli pezzi, prevenendo, quindi, la sua proliferazione. Gli enzimi che tagliano il DNA virale vennero chiamati **enzimi di restrizione**. È anche stato osservato che nei ceppi in cui il fago può proliferare il DNA virale risulta modificato tramite metilazione da una metilasi presente nel batterio e ciò impedisce l'azione degli enzimi di restrizione sul DNA virale. In sostanza, il sistema di modificazione e restrizione dell'ospite è utilizzato dai batteri

per difendersi dalle infezioni virali: il DNA del virus che entra nel batterio non è metilato e, perciò, viene attaccato e degradato dagli enzimi di restrizione; al contrario, il DNA della cellula ospite, che è metilato da enzimi chiamati **metilasi** e che modificano il DNA batterico a ogni ciclo di replicazione del DNA, non è attaccabile dagli enzimi di restrizione stessi. La purificazione di centinaia di enzimi di restrizione da numerose specie batteriche ha permesso la produzione di reagenti molecolari essenziali per lo sviluppo delle tecnologie del DNA ricombinante.

Gli enzimi di restrizione prendono il nome dall'organismo dal quale vengono isolati. Per convenzione si usa un sistema a tre lettere in corsivo. Se un ospite particolare ha diversi sistemi di restrizione-modificazione, questi vengono identificati con i numeri romani (ad es. *HaeI*, *HaeII*, *HaeIII*). Le lettere aggiuntive indicano un particolare ceppo batterico da cui è stato ottenuto l'enzima: per esempio, *BglII* viene da *Bacillus globigii*; *EcoRI* deriva dal ceppo di *Escherichia coli* RY13 ecc.

Gli ER vengono distinti in ER del I e del II tipo. Gli **ER del I tipo** riconoscono una specifica sequenza di nucleotidi, ma tagliano il DNA in posizioni non specifiche a una certa distanza dalla sequenza riconosciuta. Poiché il sito di taglio non corrisponde a una specifica coppia di basi, gli enzimi di restrizione del I tipo non sono quasi mai utilizzati nelle tecnologie del DNA ricombinante.

Gli **ER del II tipo** riconoscono anch'essi specifiche sequenze di nucleotidi, ma tagliano il DNA all'interno di quelle specifiche sequenze, chiamate **siti di taglio** (fig. 20.15A). La sequenza di riconoscimento per la maggioranza degli ER del II tipo (ne esistono circa 300) è lunga 4-6 nucleotidi, ma esistono ER chiamati **rare cutter** (tagliatori poco frequenti) che riconoscono sequenze più lunghe. È meglio chiarire subito che la frequenza di taglio di un ER è legata alla lunghezza della sequenza riconosciuta da quello specifico enzima. Infatti, la probabilità che

A Siti di riconoscimento e taglio di alcuni ER

ER che riconoscono 4 pb

MboI

5'-GATC-3'
3'-CTAG-5'

HaeII

5'-GGCC-3'
3'-CCGG-5'

AluI

5'-AGCT-3'
3'-TCGA-5'

HhaI

5'-GCGC-3'
3'-CGCG-5'

MboI

5'-GTAC-3'
3'-CATG-5'

ER che riconoscono 6 pb

EcoRI

5'-GAATTC-3'
3'-CTTAAG-5'

HindIII

5'-AAGCTT-3'
3'-TTCAAG-5'

BamHI

5'-GGATCC-3'
3'-CCTAGG-5'

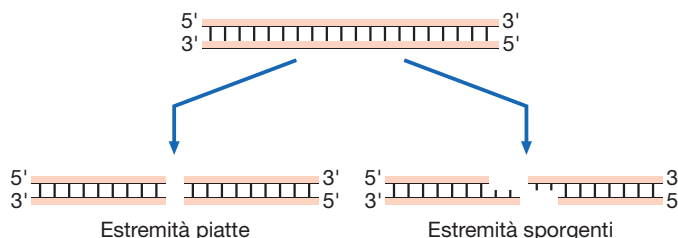
Sall

5'-GTCGAC-3'
3'-CAGCTG-5'

SmaI

5'-CCCGGG-3'
3'-GGGCCC-5'

B Gli ER possono tagliare il DNA generando estremità piatte o sporgenti



C Le estremità sporgenti prodotte dagli ER possono protrudere al 5' o al 3'

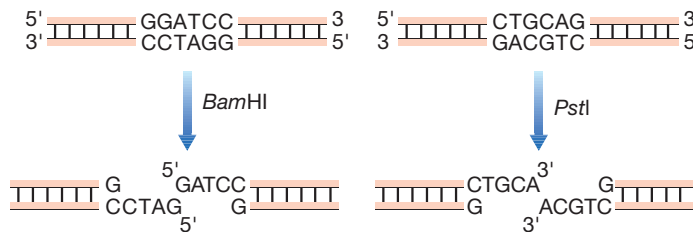


Figura 20.15 Siti di riconoscimento e modalità di taglio sul DNA da parte degli enzimi di restrizione. Gli enzimi di restrizione (ER) sono delle endonucleasi che riconoscono specifiche sequenze di nucleotidi sul DNA. La lunghezza delle sequenze riconosciute dai vari ER varia da un minimo di 4 a un massimo di 8 nucleotidi. Nel pannello (A) sono indicate le sequenze riconosciute da alcuni enzimi di restrizione e il modo in cui questi tagliano. (B) Gli ER

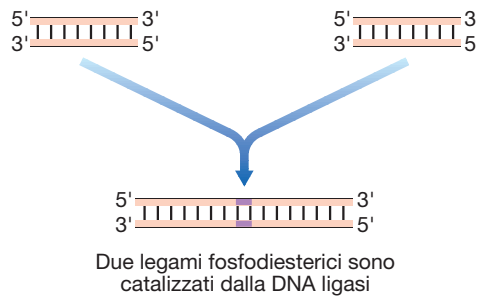
possono tagliare nel mezzo la sequenza riconosciuta generando delle estremità "piatte" o blunt; alternatively, gli ER possono tagliare la sequenza riconosciuta in modo asimmetrico sui due filamenti di DNA lasciando delle estremità "sporgenti" o protruding. (C) Queste estremità a singolo filamento possono sporgere al 5' o al 3', come mostrato per i due esempi *BamHI* e *PstI*.

sul DNA si trovi una qualsiasi, ma specifica sequenza di 4 paia di basi (pb) è di $(1/4)^4$, cioè 1 volta ogni 256 nucleotidi; se la sequenza riconosciuta è lunga 6 pb si troverà statisticamente sul DNA 1 volta ogni 4096 pb $(1/4)^6$. È chiaro che enzimi che riconoscano sequenze di 8 pb o più taglieranno il DNA molto raramente (rare cutter). ER isolati da batteri differenti ma che riconoscono e tagliano la stessa sequenza nucleotidica vengono chiamati **isoschizomeri**.

La maggior parte degli ER riconosce sequenze sulla doppia elica del DNA definite **palindromiche**: cioè hanno la stessa sequenza sui due filamenti letti in direzione 5'→3' (ad es. la sequenza 5'GAATTC3' riconosciuta dall'ER noto come *EcoRI* o la sequenza 5'GGCC3' riconosciuta da *HaeIII*).

Esistono ER in grado di tagliare entrambi i filamenti della doppia elica nello stesso punto, generando delle **estremità definite piatte** o **blunt ends**; altri ER, invece, operano un taglio obliquo lasciando delle **estremità sporgenti** (**fig. 20.15B**). In questo caso le estremità possono sporgere, a seconda del particolare enzima considerato, al 5' o al 3' e vengono chiamate **sticky ends** (estremità appiccicose o coesive) (**fig. 20.15C**). Frammenti di DNA generati da tagli con ER possono essere legati tra di loro o inseriti in un vettore di clonaggio in provetta utilizzando l'enzima DNA ligasi (**fig. 20.16**). Come descritto nel capitolo 6, la

A Ligazione di DNA



B Estremità "sporgenti" coesive vengono legate più efficacemente che estremità "piatte"

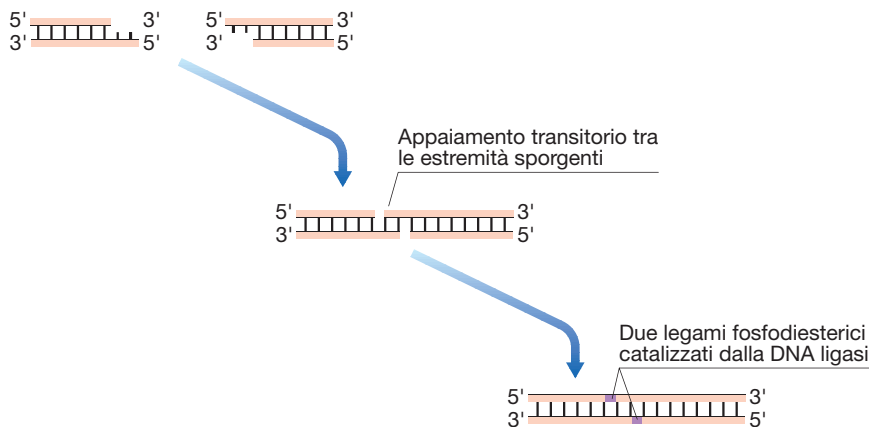


Figura 20.16 Ligazione di molecole di DNA *in vitro* con la DNA ligasi. Per legare tra di loro due molecole di DNA *in vitro*, la DNA ligasi deve catalizzare la formazione di due legami fosfodiesterici, uno per ciascun filamento di DNA. Nel pannello (A) è schematizzata la ligazione *in vitro* di due molecole di DNA con estremità piatte. Nel pannello (B) è mostrata la ligazione di due molecole di

DNA con estremità sporgenti in 5' generate dal taglio con lo stesso enzima di restrizione. La ligazione di due molecole di DNA con estremità coesive sporgenti è più efficiente che la ligazione di due molecole con estremità piatte: l'appaiamento transiente tra le due estremità sporgenti coesive facilita, infatti, l'azione della DNA ligasi nel catalizzare i due legami fosfodiesterici.

DNA ligasi gioca un ruolo essenziale durante la replicazione del DNA, ma è diventata un reagente fondamentale per costruire molecole di DNA ricombinanti in provetta. Frammenti di DNA con estremità sia piatte che sporgenti possono essere legati tra di loro dalla DNA ligasi del batteriofago T4. Frammenti che abbiano estremità sporgenti possono essere legati tra di loro solo se sono stati tagliati con lo stesso ER o con ER diversi ma che lascino le stesse estremità coesive. Frammenti che lasciano estremità coesive sono legati tra di loro più efficientemente che frammenti con estremità piatte. Va sottolineato che la ligazione di frammenti con estremità sporgenti generate dallo stesso ER ricostruisce sulla molecola di DNA ricombinante il sito di restrizione originario. Lo stesso accade se vengono legati tra di loro frammenti con estremità piatte generati con lo stesso ER; se, però, si legano tra loro frammenti di restrizione generati da ER che lasciano estremità piatte ma generati da ER che riconoscono sequenze diverse, la loro successiva saldatura non ricostruirà la sequenza nucleotidica originaria.

Vettori di clonaggio

Come anticipato precedentemente (**fig. 20.14**), il clonaggio molecolare di un frammento di DNA parte da una miscela complessa di frammenti generati dalla rottura meccanica o dal taglio con ER del genoma di un certo organismo e usa poi cellule viventi per riprodurre molte copie di uno specifico frammento che è stato inserito all'interno di "trasportatori specializzati" genericamente chiamati **vettori di clonaggio**. Esistono numerosi vettori di clonaggio e in questo capitolo ci limiteremo a descrivere i più comuni; molecole di DNA ricombinanti costruite in un dato vettore di clonaggio possono essere propagate in organismi diversi, anche se l'ospite di gran lunga più utilizzato è il batterio *Escherichia coli*.

Un vettore di clonaggio deve avere alcune caratteristiche particolari (**fig. 20.17**):

- deve contenere una origine di replicazione che gli permetta di replicarsi autonomamente una volta introdotto all'interno della cellula ospite. Vettori privi di un'origine di replicazione possono essere utilizzati in casi particolari in cui si vuole integrare del DNA esogeno all'interno del genoma della cellula ospite, magari in precise posizioni (vedremo più avanti tali applicazioni);
- deve contenere uno o più "marcatori di selezione" che permettano di individuare le cellule che contengono il vettore e qualsiasi frammento di DNA in esso inserito;
- deve contenere il maggior numero possibile di siti di restrizione unici, cioè presenti una sola volta nel loro DNA: ciò permette di poter inserire all'interno del vettore una varietà di frammenti di restrizione.

Plasmidi

I plasmidi sono i vettori di clonaggio più frequentemente utilizzati nelle tecnologie del DNA ricombinante. Essi sono degli elementi genetici costituiti da molecole di DNA circolare a doppio filamento piuttosto piccole (da 2 a 5 kpb) e i più utilizzati sono in grado di replicarsi autonomamente all'interno di cellule di *E. coli*. Molti dei plasmidi presenti in natura contengono già dei marcatori selettivi che conferiscono alle cellule batteriche che contengono tali plasmidi la resistenza ad alcune sostanze, per esempio a degli antibiotici. Per di più, molti plasmidi esistono all'interno delle cellule batteriche in numerose copie. I plasmidi naturali, nel corso degli anni, sono però stati modificati dai ricercatori così da far loro acquisire caratteristiche addizionali che li rendono sempre più versatili per le svariate applicazioni delle tecnologie del DNA ricombinante.

In **figura 20.17** è mostrato uno schema del **plasmide pBR322**, che è il prototipo da cui sono stati derivati numerosi altri plasmidi utilizzati in tecniche di clonaggio molecolare. Il plasmide pBR322 contiene un'origine di replicazione che gli

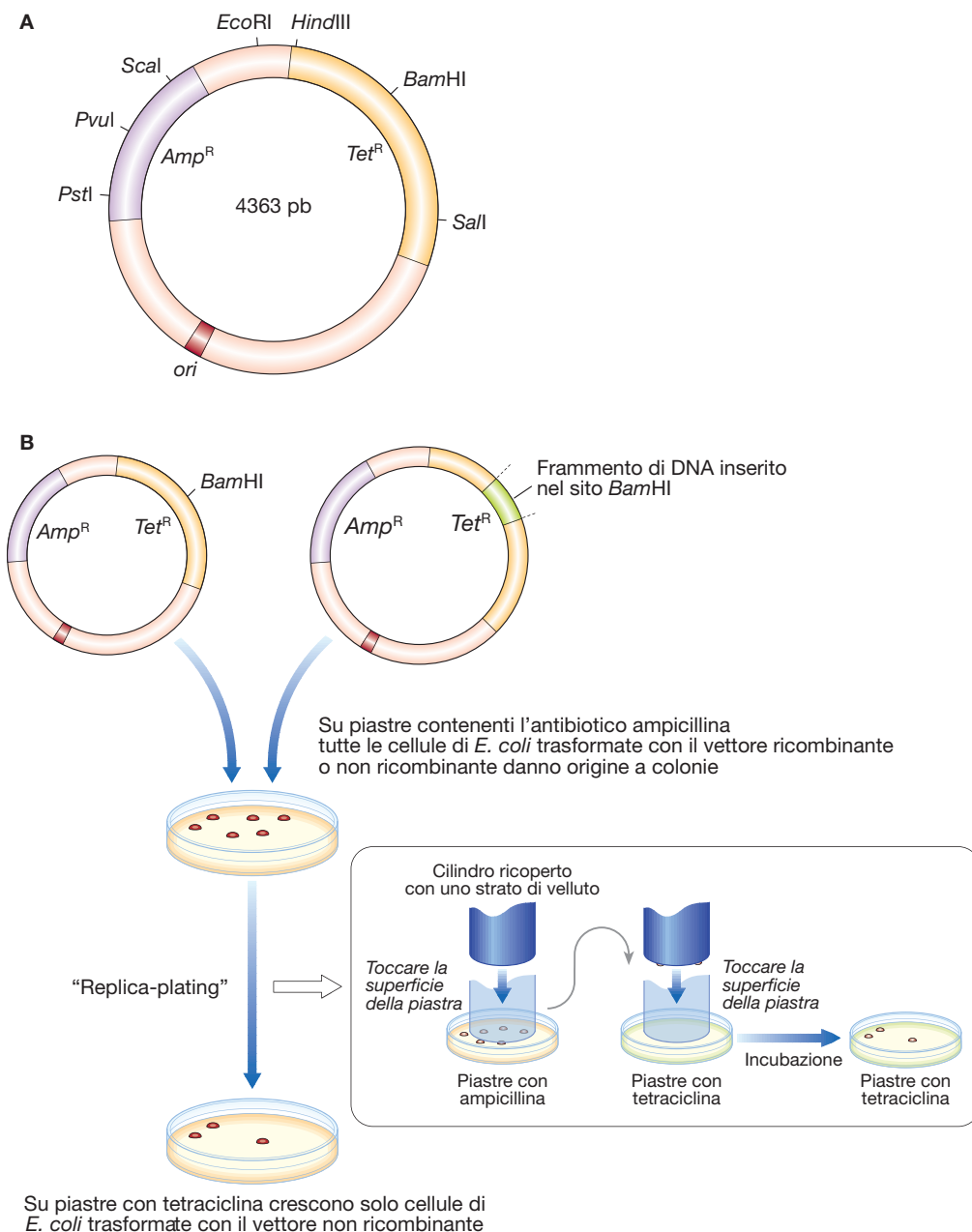


Figura 20.17 Plasmide pBR322 come prototipo di un vettore di clonaggio e utilizzo di geni che conferiscono la resistenza ad antibiotici come marcatori di selezione di molecole di DNA ricombinanti. (A) Struttura del plasmide pBR322. La mappa mostra la posizione dell'origine di replicazione del plasmide (*ori*) e quella dei geni che conferiscono la resistenza agli antibiotici ampicillina (*Amp^R*) e tetraciclina (*Tet^R*). Sono indicate anche le posizioni dei siti di riconoscimento unici per alcuni ER. (B) Dopo il clonaggio *in vitro* possono essersi formate molecole di DNA ricombinanti con inserito un frammento di DNA esogeno (a destra) o il vettore può essersi richiuso su se stesso, ricostituendo un vettore di clonaggio non ricombinante (a sinistra). Dopo trasformazione, le cellule di *E. coli* che hanno acquisito un vettore ricombinante o non ricombi-

nante vengono selezionate su terreno contenente l'antibiotico ampicillina. Le colonie ottenute sono poi in parte trasferite mediante la tecnica del replica-planting (vedi finestra 6.2) su piastre contenenti l'antibiotico tetraciclina. Ricordando che l'inserimento di un frammento di DNA esogeno nel sito *Bam*HI del vettore di clonaggio inattiva il gene che conferisce resistenza alla tetraciclina (*Tet^R*), ne deriva che cellule trasformate con plasmidi ricombinanti non saranno in grado di formare colonia su terreno solido contenente tale antibiotico. Le colonie che sono state trasformate da plasmidi ricombinanti saranno, quindi, quelle in grado di formare colonie su terreno addizionato con ampicillina, ma non con tetraciclina. Si può risalire a tali colonie riferendosi alle colonie che crescevano sulla prima piastra (piastra madre).

permette di replicarsi autonomamente in cellule di *E. coli*, due geni (chiamati *Amp^R* e *Tet^R*) che consentono alle cellule batteriche che contengono il plasmide di poter crescere in presenza degli antibiotici ampicillina e tetraciclina. Nello schema di **figura 20.17** sono inoltre indicati siti riconosciuti da specifici ER che sono presenti una sola volta nel DNA plasmidico e, quindi, utili per esperimenti di clonaggio.

Analizziamo come si procede per clonare frammenti di DNA all'interno del vettore pBR322: come vedremo appresso, un utile sito di restrizione per clonare frammenti di DNA all'interno del plasmide pBR322 è il sito *Bam*HI presente all'interno del gene *Tet^R* che conferisce la resistenza all'antibiotico tetraciclina. *Bam*HI è un ER che genera delle estremità coesive sporgenti in 5' e, quindi, dopo taglio con *Bam*HI del vettore e l'aggiunta di frammenti *Bam*HI di DNA proveniente da qualsiasi organismo, la DNA ligasi, con una certa probabilità, sarà in grado di inserire un frammento *Bam*HI esogeno all'interno del vettore, generando così una molecola di DNA ricombinante. Nella reazione *in vitro* con la DNA ligasi possono formarsi diversi prodotti, ma quelli che in questo contesto sono più rilevanti sono due: la ligasi può richiudere su se stesso il vettore tagliato con *Bam*HI o può inserire al suo interno il frammento di DNA esogeno. Nel primo caso si ricostituisce una normale molecola di pBR322 con i due geni che conferiscono la resistenza all'ampicillina e alla tetraciclina. Nel secondo caso, il gene *Tet^R* viene interrotto dall'inserimento del frammento esogeno e, quindi, la sua capacità funzionale di conferire resistenza alla tetraciclina verrà abolita. A questo punto si devono introdurre le molecole ricombinanti costruite *in vitro* all'interno dell'ospite in cui possono propagarsi, cioè in cellule di *E. coli*.

Trasformazione di cellule di *Escherichia coli* e selezione dei trasformanti

Si è già appreso nel corso di Genetica che è possibile introdurre molecole di DNA esogene all'interno di batteri attraverso un processo chiamato **trasformazione**. Alcuni batteri sono naturalmente capaci di far entrare al loro interno molecole di DNA esogene: per esempio il batterio *Streptococcus pneumoniae*, che è stato utilizzato da Griffith e collaboratori per identificare il principio trasformante descritto nel paragrafo 1.3. Le cellule di *E. coli*, tuttavia, non sono naturalmente trasformabili; possono, però, acquisire questa capacità dopo trattamenti molto semplici che alterano la permeabilità della membrana dei batteri senza affliggere la loro vitalità. Ciò può essere ottenuto mediante **trasformazione chimica** che si basa sul trattamento dei batteri con CaCl_2 , oppure mediante **elettroporazione**, che consiste in una breve e intensa scarica elettrica che provoca l'apertura di pori nella membrana.

Per proseguire nell'esperimento di clonaggio che stiamo descrivendo, dobbiamo quindi, per prima cosa, rendere trasformabili (spesso si dice "competenti" alla trasformazione) le cellule di *E. coli*. Le cellule competenti ottenute sono mescolate alla miscela della reazione di clonaggio *in vitro* descritta poco sopra e, dopo aver atteso un po' di tempo per permettere l'eventuale entrata del DNA nelle cellule batteriche, queste ultime sono piastrate su capsule Petri contenenti terreno solido addizionato con l'antibiotico ampicillina. Tutte le cellule batteriche prive del plasmide erano originariamente sensibili all'ampicillina: quindi, le cellule batteriche che non sono state in grado d'introdurre alcune molecole di DNA al loro interno (sono la grande maggioranza) non saranno in grado di moltiplicarsi e, quindi, di formare colonie sul terreno solido utilizzato. Cresceranno e formeranno colonie solo le cellule batteriche che al loro interno portano il vettore che nella reazione di ligazione si è richiuso su se stesso o i plasmidi ricombinanti che hanno integrato al loro interno il frammento di DNA esogeno. Infatti, in entrambi i casi, il gene *Amp^R* che conferisce la resistenza all'ampicillina non è stato intaccato dalla tecnica di clonaggio. È possibile, però, distinguere le colonie che hanno al loro interno il plasmide ricombinante facendo un semplice esperimento di replica-plating che

abbiamo già descritto nel capitolo 6. Come mostrato in **figura 20.17**, le colonie batteriche che hanno introdotto il frammento di DNA esogeno nel sito *Bam*HI del vettore di clonaggio hanno perso la funzionalità di *Tet*^R e, quindi, non saranno capaci di formare colonie dopo essere state trasferite su un terreno che contiene l'antibiotico tetraciclina. Individuate tali colonie sarà possibile risalire sulla prima piastra (piastra madre addizionata con ampicillina) alla posizione delle colonie in grado di formare colonia su ampicillina, ma non su tetraciclina: tali colonie dovrebbero contenere al loro interno il plasmide ricombinante.

È però necessario confermare sperimentalmente tale possibilità: a tal fine, le cellule di *E. coli* capaci di crescere su terreno con ampicillina, ma non su terreno con tetraciclina, vengono fatte ulteriormente crescere e da tali cellule si estrae il DNA plasmidico. È atteso che il plasmide pBR322 ricombinante sia più grande del plasmide originario, dato che contiene al suo interno un frammento di DNA esogeno che ne aumenta la dimensione: quindi, i plasmidi ricombinanti possono essere distinti dal plasmide originario mediante elettroforesi del DNA su gel (**finestra 20.3**).

Plasmidi della serie pUC: selezione bianco o blu

Nel corso degli anni sono stati sviluppati vettori plasmidici via via più sofisticati e che utilizzano, per esempio, sistemi di selezione alternativi. Tra questi, il **vettore di clonaggio pUC19** (**fig. 20.18A**) ha inserito al suo interno, oltre al gene che conferisce la resistenza all'antibiotico ampicillina, un segmento di DNA sintetizzato *in vitro*, chiamato **MCS** (*Multiple Cloning Site*). Come indicato in figura, il segmento MCS contiene condensati i siti di restrizione riconosciuti da 11 ER diversi in circa 50 nucleotidi. Inoltre, il segmento MCS si trova all'interno del gene *lacZ*'5' di *E. coli*. Il gene *lacZ* codifica per l'enzima β -galattosidasi: cellule batteriche che producono un enzima attivo danno origine a delle colonie di colore blu se cresciute su un terreno di crescita contenente un substrato cromogeno della β -galattosidasi, chiamato X-gal. Sul plasmide pUC19 non è presente l'intero gene *lacZ*, ma la porzione in 5' del gene, codificante per la porzione N-terminale della proteina. Nel genoma delle cellule di *E. coli* che si utilizzano per la trasformazione è presente la porzione in 3' del gene *lacZ* (*lacZ*'3') codificante per la porzione C-terminale della β -galattosidasi. Questa porzione di proteina non è funzionante: tuttavia, se nella stessa cellula viene prodotta contemporaneamente anche la sua porzione N-terminale codificata dal segmento di gene presente sul plasmide pUC19, attraverso un fenomeno chiamato **α -complementazione** nella cellula riesce ad assemblarsi un enzima attivo in grado di metabolizzare il substrato X-gal e tali cellule daranno origine a colonie di colore blu. L'inserzione del segmento MCS nella porzione di gene *lacZ*'5' non modifica il modulo di lettura del gene e aggiunge solo alcuni amminoacidi in più che non alterano l'attività della β -galattosidasi dopo α -complementazione. Se, però, come schematizzato in **figura 20.18B**, si ottiene nella procedura di clonaggio un plasmide ricombinante inserendo in un sito di restrizione presente nell'MCS un frammento di DNA esogeno, il gene *lacZ*'5' viene inattivato e, quindi, dopo trasformazione delle cellule batteriche non verrà prodotta β -galattosidasi attiva. In pratica, le cellule di *E. coli* trasformate sono selezionate per la loro capacità di crescere su terreno contenente l'antibiotico ampicillina: le cellule contenenti il vettore non ricombinante formeranno colonie di colore blu in quanto in grado di metabolizzare X-gal dopo α -complementazione, mentre le cellule batteriche contenenti il plasmide ricombinante, essendo incapaci di produrre un enzima attivo, daranno origine a colonie di colore bianco.

Vettori di espressione e produzione di proteine ricombinanti

Negli esempi descritti sopra, segmenti di DNA sono clonati all'interno di un vettore plasmidico che può essere propagato in cellule di *E. coli* così da poter ottenere

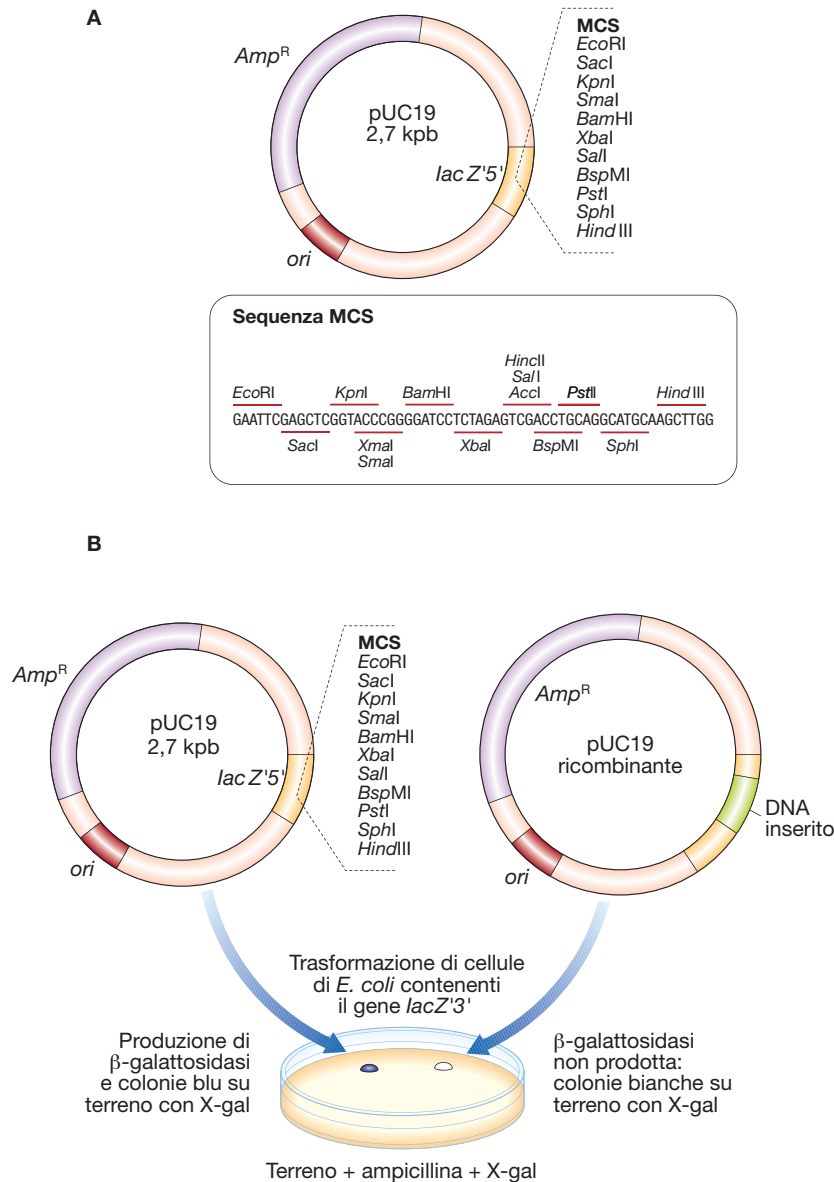


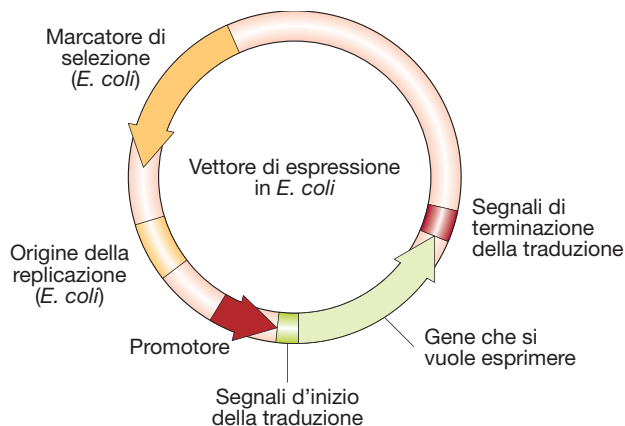
Figura 20.18 Clonaggio in plasmidi della serie pUC: selezione con la tecnica del “bianco o blu”. (A) Il vettore pUC19 è un tipico vettore per la selezione “biancoblu”. Il gene *Amp^R* conferisce alle cellule trasformate con il vettore ricombinante o non ricombinante la capacità di crescere su terreno contenente l’antibiotico ampicillina. La sequenza MCS contiene al suo interno le sequenze

nucleotidiche riconosciute dagli ER indicati. *Ori* rappresenta l’origine di replicazione del plasmide, mentre la funzione di *lacZ'5'* è descritta nel testo. (B) Nel pannello è descritto lo schema di un esperimento di clonaggio nel vettore pUC19: i sistemi di selezione e il riconoscimento dei trasformanti tramite il colore delle colonie sono descritti ampiamente nel testo.

grandi quantità di quel frammento di DNA, non comportando però l’espressione di geni in esso eventualmente presenti. I **vettori d’espressione**, invece, sono una categoria speciale di vettori all’interno dei quali si introduce un frammento di DNA contenente un gene codificante per una specifica proteina che si intende fare esprimere nelle cellule trasformate per produrne grandi quantità.

Il gene che si vuole esprimere proviene, di solito, da un organismo diverso da quello utilizzato per il clonaggio: per esempio, si può esprimere un gene umano all’interno di cellule di *E. coli*. Le proteine prodotte tramite l’utilizzo di vettori di espressione sono spesso indicate come **proteine ricombinanti**. La sintesi di proteine ricombinanti in *E. coli* è molto utilizzata, in quanto si raggiungono i livelli

Figura 20.19 Schema di un vettore plasmidico per l'espressione di proteine ricombinanti in *Escherichia coli*. Nello schema sono indicati solo gli elementi importanti per il mantenimento del plasmide, la sua selezione e l'espressione del gene che si vuole esprimere.



elevati di produzione di proteina richiesti per studi strutturali, per la produzione di anticorpi e per una varietà di altri utilizzi biotecnologici. Basti pensare che la produzione di proteine ricombinanti da parte di imprese biotecnologiche ha prodotto nel 2008 un fatturato globale di oltre 70 miliardi di dollari.

Proteine ricombinanti eterologhe vengono prodotte in diversi organismi, quali lievito, cellule d'insetto, sistemi virali, cellule di mammifero in coltura ecc.; tuttavia, l'espressione in cellule di *E. coli* è di gran lunga la più semplice ed economica. Quando si vuole esprimere uno specifico gene bisogna, innanzitutto, decidere quale ospite si vuole utilizzare per l'espressione e scegliere il vettore più appropriato. Esistono decine di diversi vettori di espressione per la produzione di proteine eterologhe in vari organismi. Qui indicheremo soltanto le caratteristiche generali di un vettore di espressione in *E. coli*, lasciando a testi specializzati una descrizione più completa delle varie possibilità e utilizzi.

I meccanismi alla base del controllo dell'espressione genica nei procarioti sono, come abbiamo visto, molto diversi da quelli che controllano l'espressione di un gene eucariotico. Ne deriva che, se vogliamo esprimere un gene umano per produrre la corrispondente proteina in cellule di *E. coli*, dobbiamo clonare tale gene in un vettore che utilizzi i sistemi di regolazione del batterio.

Un tipico vettore di espressione in *E. coli* è rappresentato in **figura 20.19** e contiene i seguenti elementi fondamentali:

- un'origine di replicazione;
- un marker per la selezione dei trasformanti contenenti il vettore ricombinante;
- un promotore espresso in cellule di *E. coli* con i segnali regolativi richiesti per ottimizzare la produzione di mRNA;
- segnali d'inizio e di terminazione della traduzione.

Altri "espedienti" genetici sono utilizzati per specifiche applicazioni o per facilitare la purificazione della proteina prodotta come vedremo appresso.

La maggior parte dei promotori utilizzati per l'espressione di una proteina eterologa, sia in *E. coli* che in altri organismi, sono regolabili. Infatti, la proteina eterologa prodotta tende a essere identificata come "estranea" dalla cellula ospite e, spesso, viene degradata da specifiche proteasi. Alcune proteine, soprattutto se prodotte nelle grandi quantità ottenibili con i vettori di espressione, possono anche essere tossiche per le cellule o, comunque, interferire con la loro velocità di crescita. La possibilità di indurre sperimentalmente la produzione di una proteina in specifiche condizioni permette, da una parte, di non interferire con la crescita delle cellule fino al momento dell'induzione e, dall'altra, di verificare l'efficienza d'espressione paragonando i livelli della proteina in questione prima e dopo l'induzione dell'espressione del gene d'interesse.

Tabella 20.2 Produzione di proteine eterologhe in *Escherichia coli*.

Vantaggi
• Vasta scelta di vettori di clonaggio
• Vasta scelta di ceppi batterici
• Controllo relativamente semplice dell'espressione del gene
• Ottima resa nella produzione di proteine ricombinanti (25% del totale)
• La proteina ricombinante può essere espressa come proteina di fusione
• La proteina ricombinante può essere secreta nel terreno di crescita
• Basso costo di produzione
• Processi di fermentazione ben conosciuti
• Facile possibilità di "scale-up"
Svantaggi
• La proteina ricombinante può mancare di modificazioni post-traduzionali (ad es. glicosilazioni)
• Batteri Gram-negativi contengono endotossine
• Presenza di proteasi
• La sovrapproduzione di proteine può portare alla formazione di "corpi d'inclusione", aggregati insolubili che rendono difficoltosa la purificazione della proteina e possono ridurre la sua attività biologica

Con specifici vettori di espressione la proteina prodotta in *E. coli* può raggiungere anche il 25% delle proteine totali presenti nella cellula: nella maggior parte dei casi, però, tale elevato livello d'espressione fa sì che la proteina prodotta non sia solubile, ma venga inclusa in aggregati, chiamati **corpi d'inclusione**, che possono essere solubilizzati solo in condizioni sperimentali piuttosto drastiche. La produzione di proteine eterologhe in batteri ha quindi una serie di vantaggi, ma anche di svantaggi, che sono riassunti in **tabella 20.2**.

Lo scopo finale della produzione di proteine in sistemi eterologhi (siano essi cellule di *E. coli*, di lievito, di mammifero o di altri organismi) consiste, generalmente, nel purificare la proteina in questione. A tal fine si utilizzano una serie di "espedienti" per facilitare la purificazione della proteina in esame. Il trucco più frequentemente utilizzato è quello di fondere il gene che deve essere espresso con sequenze codificanti per peptidi, chiamate **protein tag** (o etichette), determinando la produzione di una proteina modificata che può però essere facilmente purificata o identificata in estratti proteici analizzati tramite la tecnica del Western blot (**finestra 20.4**). Esistono diversi tipi di tag, ma il loro utilizzo è sostanzialmente legato al fatto che il peptide codificato dal tag permette il legame della proteina modificata a specifici ligandi.

La modificazione della proteina d'interesse con un "tag" contenente 6 residui di istidina è uno dei sistemi più utilizzati. Come si può osservare in **figura 20.20**, la proteina ricombinante *His-tagged* ha una particolare affinità per una resina a cui è legato acido nitriltriacetico (Nta) complessato con ioni nichel bivalenti. Una proteina contenente "His-tag" si lega specificamente alla resina, tramite l'interazione tra i residui di istidina e gli ioni nichel, mentre proteine prive del "tag" non rimangono attaccate alla resina. L'eluizione della proteina-tagged in specifiche condizioni può portare a una sostanziale purificazione della proteina ricombinante in un unico passaggio.

Sono stati sviluppati una varietà di analoghi "trucchi" per facilitare la purificazione di proteine-tagged ricombinanti; va, comunque, sottolineato che una proteina-tagged è diversa dalla proteina nativa. Se, quindi, si vuole valutare la

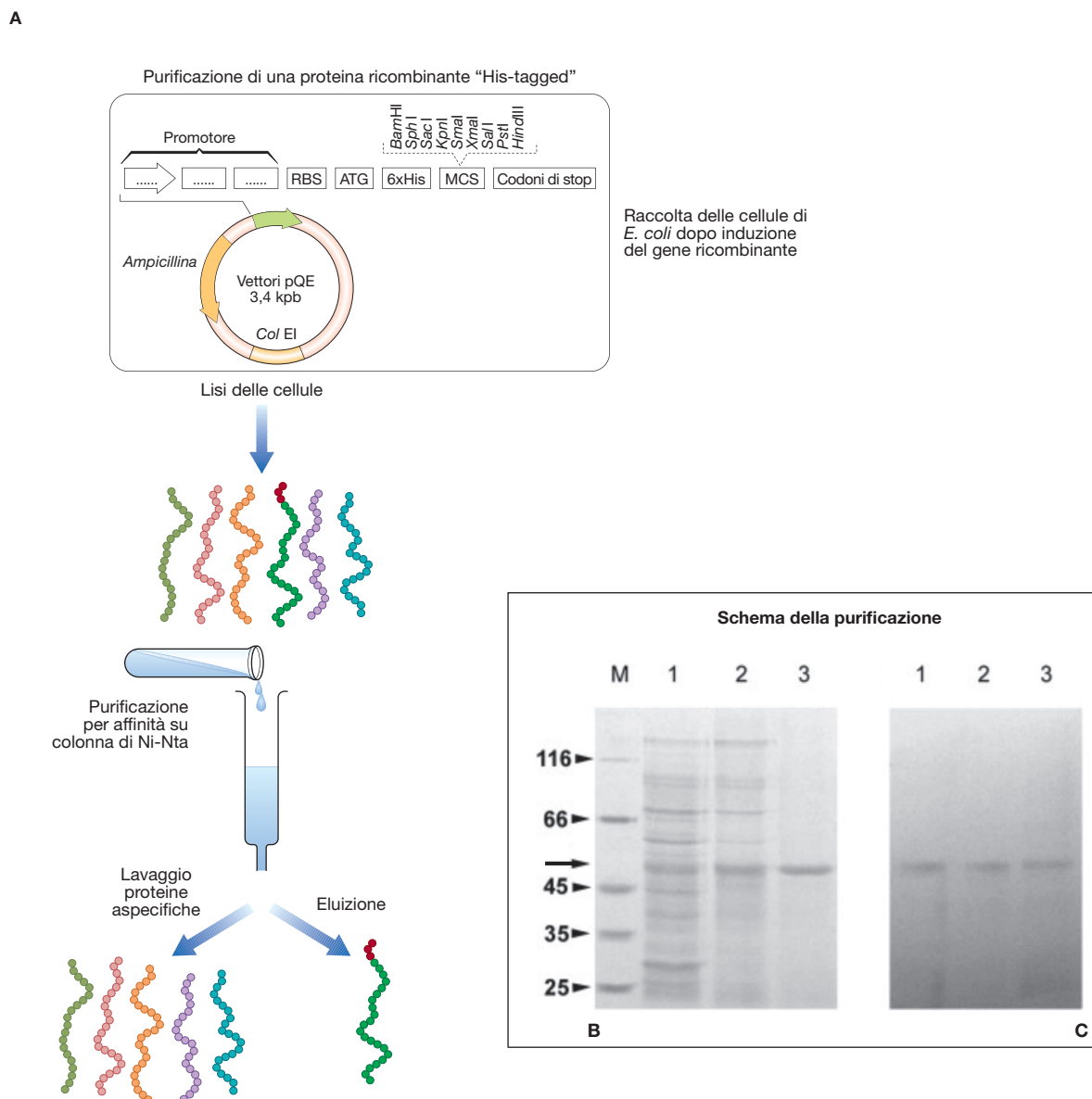


Figura 20.20 Schema della purificazione di una proteina ricombinante marcata con il "6-His-Tag" prodotta in cellule di *E. coli*. **(A)** Schema della produzione e purificazione della proteina ricombinante. Nella parte alta del pannello è indicato uno schema di un vettore d'espressione della serie pQE. Con il termine "promotore" vengono indicate delle regioni sul plasmide che controllano l'espressione del gene clonato a livello trascrizionale. RBS: sito di legame dei ribosomi richiesti per la traduzione dell'mRNA prodotto. ATG: tripletta d'inizio della traduzione. 6xHis: tag contenente 6 residui d'istidina. MCS: segmento di clonaggio multiplo seguito da codoni di terminazione della traduzione. Dopo lisi delle cellule batteriche, la proteina ricombinante (in verde scuro) con il tag di 6 istidine indicato in rosso, si può legare alla resina con gli ioni Ni^{++} e, dopo lavaggio della resina, la proteina ricombinante può essere specifica-

mente eluita. **(B)** Schema della purificazione. In corsia M sono stati separati, tramite elettroforesi SDS-PAGE, dei marker di proteine a peso molecolare noto. In corsia 1 sono state separate proteine presenti in un estratto di cellule di *E. coli* trasformate con il vettore ricombinante indicato nel pannello A che esprime una proteina-His-tag di circa 50 kDa (indicata dalla freccia). Dopo il caricamento dell'estratto cellulare sulla resina contenente gli ioni nichel, la resina è lavata per eliminare il materiale non fortemente legato (analizzato in corsia 2), mentre le proteine legate alla resina e poi opportunamente eluite sono analizzate in corsia 3. Le proteine separate nelle quattro corsie sono state colorate con Coomassie blue. **(C)** Western blot (vedi finestra 20.4) delle proteine separate nelle corsie 1-3 mostrate nel pannello B dopo incubazione con un anticorpo che riconosce specificamente la proteina X.

funzione della proteina così modificata all'interno di una cellula è necessario assicurarsi preventivamente che la proteina modificata abbia mantenuto le proprietà funzionali della proteina naturale. Anche per molti studi strutturali o biochimici accurati *in vitro* è necessario disporre della proteina naturale. A tal fine sono state sviluppate delle procedure che utilizzano uno o più "tag" durante la fase di

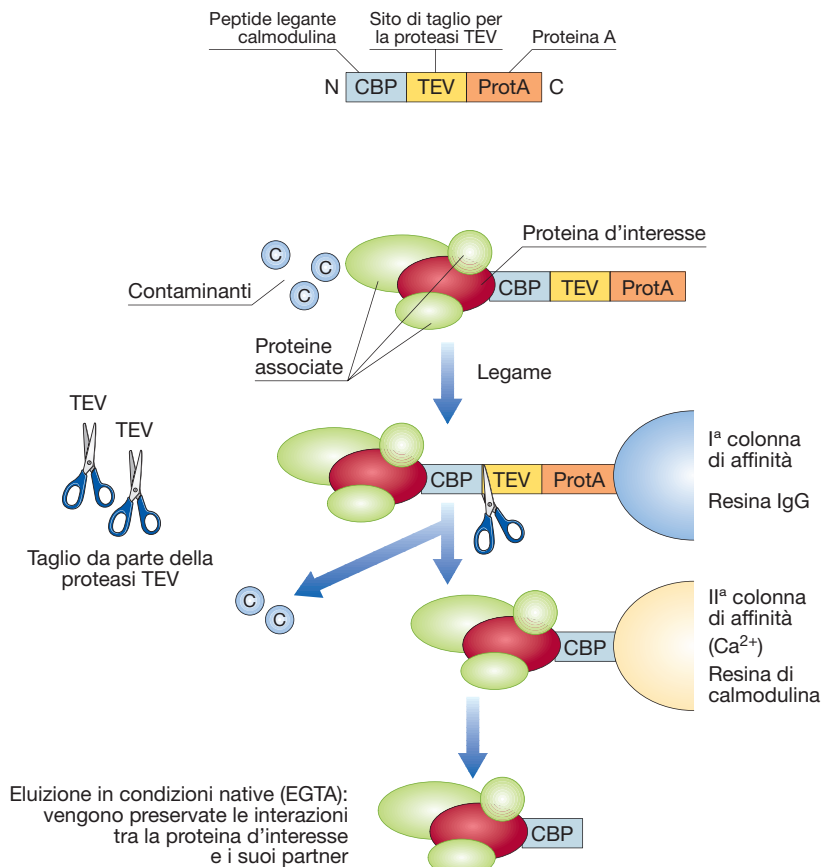
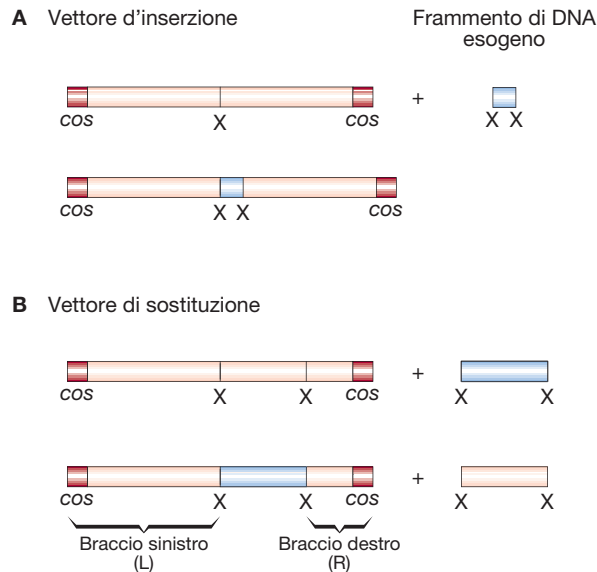


Figura 20.21 Schema dei vari passaggi di purificazione utilizzando la strategia nota come TAP: *Tandem Affinity Purification*. La procedura TAP è stata con successo utilizzata per purificare non solo la proteina d'interesse, ma anche altre proteine che con quella interagiscono, permettendo la purificazione d'interi complessi proteici esistenti nella cellula.

purificazione che vengono poi eliminati durante la procedura, così da ottenere una proteina purificata non modificata.

Un'applicazione raffinata di tale strategia è la procedura nota come **Tandem Affinity Purification** o **TAP-tagging** (fig. 20.21). Tale tecnologia permette non solo la purificazione della proteina d'interesse, ma anche l'isolamento di altri fattori proteici che interagiscono fisicamente con la proteina in esame. Brevemente, il gene che codifica per la proteina è fuso a un segmento di DNA codificante sia per un peptide in grado di legare la calmodulina (*Calmodulin Binding Peptide*, CBP) che per la proteina A, un polipeptide prodotto da cellule batteriche che ha un'estrema affinità per una porzione delle immunoglobuline (IgG) di mammifero. Questi due domini proteici sono fra loro separati da una serie di amminoacidi che sono il bersaglio della proteasi del virus del tabacco (TEV). L'espressione di un tale costrutto porta alla produzione di una proteina di fusione in grado di mantenere possibili legami con i suoi interattori fisiologici. La proteina di fusione viene dapprima purificata sfruttando l'interazione del dominio corrispondente alla proteina A con IgG covalentemente legate a una resina inerte. Si procede poi al taglio della proteina ricombinante legata alla resina utilizzando la proteasi TEV prodotta commercialmente. La presenza del dominio corrispondente al *Calmodulin Binding Peptide* permette un secondo passaggio di purificazione per affinità utilizzando una resina con legata la calmodulina in presenza di ioni calcio. La proteina ricombinante e i suoi eventuali interattori sono poi rilasciati dalla colonna di affinità in presenza di EGTA, un composto che è in grado di chelare gli ioni calcio interferendo, così, con il legame tra CBP e calmodulina. Gli eventuali interattori vengono, infine, identificati mediante elettroforesi e successiva analisi mediante spettrometria di

Figura 20.22 Schema dell'utilizzo del DNA del batteriofago lambda come vettore di clonaggio. (A) Schema di un vettore λ del tipo "inserzione": *cos* indica le estremità coesive del DNA lineare del batteriofago λ ; X indica un qualsiasi sito di restrizione unico nel DNA del fago in cui inserire un frammento di DNA esogeno generato con il taglio dello stesso enzima di restrizione. (B) Schema di un vettore λ del tipo "sostituzione": la regione tra i due siti di restrizione X presente nel vettore non contiene geni essenziali per l'infezione di cellule batteriche da parte del batteriofago e può essere sostituita da un frammento X esogeno in grado di ripristinare la dimensione totale del DNA del fago.



massa (*vedi* oltre). Tale procedura ha portato all'identificazione e caratterizzazione di interi complessi proteici in diversi sistemi cellulari.

Batteriofago lambda come vettore di clonaggio

Sin dai primi sviluppi delle tecnologie del DNA ricombinante, il DNA del batteriofago lambda (λ) e suoi derivati sono stati utilizzati come vettori di clonaggio. La biologia e le caratteristiche principali del DNA di lambda sono già state descritte nei capitoli 4 e 8. Il genoma di lambda di 48,5 kpb contiene numerosi siti riconosciuti dalla maggior parte degli enzimi di restrizione e, pertanto, il suo DNA selvatico non si presta a fungere da vettore di clonaggio. Per tale motivo, il DNA originale è stato modificato eliminando una serie di siti riconosciuti da enzimi di restrizione così da costruire dei derivati del fago lambda utilizzabili come vettori. I diversi vettori di lambda ricadono in due categorie principali, indicate in **figura 20.22**: i **vettori d'inserzione**, che possiedono un sito di restrizione unico in cui inserire un frammento di DNA, e i **vettori di sostituzione**, in cui una porzione del DNA di lambda è sostituita da un segmento di DNA esogeno. La regione che viene rimossa dal DNA di lambda non è necessaria per l'infettività del fago né per il ciclo litico, in quanto contiene solo geni implicati nel ciclo lisogenico che non solo non è richiesto durante gli esperimenti di clonaggio ma, anzi, potrebbe creare delle complicazioni. Il fago lambda presenta rispetto a un plasmide sia vantaggi che svantaggi: il vantaggio principale consiste nel fatto che è possibile ottenere l'impacchettamento del DNA del fago sia selvatico che ricombinante all'interno del capsido mediante una reazione *in vitro*. La disponibilità di un batteriofago infettivo aumenta di diversi ordini di grandezza l'efficienza con cui molecole di DNA ricombinanti vengono introdotte all'interno di cellule di *E. coli* rispetto alle procedure di trasformazione batterica. Inoltre, il sistema d'impacchettamento *in vitro* si basa sul meccanismo "a testa piena", per cui solo molecole di dimensioni corrispondenti a quelle del DNA del fago selvatico (circa 50 kpb) e che abbiano delle sequenze *cos* alle estremità riescono a essere inglobate con efficienza nel capsido. Ciò permette (con l'utilizzo dei vettori di sostituzione) la selezione di molecole di DNA ricombinante che introducano, tra il braccio destro e il braccio sinistro del vettore, molecole di DNA esogeno in grado di ristabilire la dimensione attesa del DNA di lambda. In genere, le dimensioni di tali frammenti di DNA sono più grandi di quelle normalmente clonabili all'interno di vettori plasmidici.

Come nel caso dei vettori plasmidici, anche i vettori di lambda sono stati via via modificati con il passare del tempo per migliorarne le prestazioni. Senza entrare in ulteriori dettagli, sono stati sviluppati vettori di lambda in cui sia possibile:

- utilizzare sistemi di selezione positiva dei fagi ricombinanti;
- permettere la trascrizione del frammento inserito nel vettore;
- esprimere proteine di fusione codificate in parte sul vettore di lambda e in parte da sequenze codificanti presente nel frammento inserito nel vettore.

Uno dei vettori più utilizzati per l'espressione di proteine di fusione è chiamato **λgt11** e il suo utilizzo per il clonaggio di geni codificanti per proteine d'interesse verrà descritto nelle pagine seguenti.

Cosmidi

Come già precedentemente accennato, molecole di DNA di lambda possono essere efficacemente impacchettate nel capsido solo se una quantità di DNA di 45-50 kpb si trova tra le due sequenze terminali *cos*. Sono stati, pertanto, costruiti dei plasmidi che, oltre ai tipici marcatori selettivi e a regioni di clonaggio contenenti siti per enzimi di restrizione, contenessero un frammento del DNA di lambda con il sito *cos*. Tali vettori di clonaggio sono chiamati **cosmidi** perché, utilizzando i siti *cos*, possono venire impacchettati all'interno di un capsido virale e, dopo infezione di cellule di *E. coli*, possono essere selezionati e mantenuti all'interno delle cellule infettate come dei plasmidi (**fig. 20.23**). Utilizzando un cosmide delle dimensioni iniziali di 5 kpb, per il meccanismo "a testa piena" funzionante nell'impacchettamento del DNA all'interno del capsido di lambda si possono inserire nei siti di clonaggio del cosmide frammenti di 40-45 kpb perché, in tal modo, si ripristina la dimensione finale di 45-50 kpb richiesta per l'impacchettamento *in vitro* del DNA nel capsido. Le particelle fagiche ottenute sono poi utilizzate per infettare cellule di *E. coli*. Il DNA del cosmide ricombinante, una volta entrato all'interno della cellula, circolarizza come fa normalmente il DNA di lambda con le estremità coesive (*vedi* cap. 8). A differenza del fago, però, il cosmide ricombinante si replica nel batterio come un plasmide e può essere selezionato utilizzando il marcatore di resistenza a un antibiotico presente sul cosmide originario.

I cosmidi sono stati i primi vettori di clonaggio che hanno permesso la costruzione di molecole di DNA ricombinanti contenenti inserti di DNA superiori a 30 kpb.

Vettori per il clonaggio e l'espressione in cellule eucariotiche

La maggior parte delle operazioni iniziali di clonaggio sono eseguite in cellule batteriche. Se, però, si sta lavorando su geni di eucarioti, è spesso importante verificare la funzione del gene clonato nell'organismo in cui il gene svolge la sua funzione fisiologica. Analogamente, è spesso importante esprimere il gene in esame nell'organismo appropriato per verificare l'effetto della sua espressione o per ottenere la produzione di un prodotto genico corretto: per esempio, con le previste modificazioni post-traduzionali.

Numerose tecniche sono state sviluppate per trasferire DNA all'interno di cellule eucariotiche. In questo paragrafo ci limiteremo a prendere in considerazione alcuni esempi che riguardano l'introduzione di DNA esogeno all'interno di cellule somatiche in coltura. Le tecniche variano a seconda che si tratti di cellule di eucarioti unicellulari (ad es. lievito), cellule animali o cellule vegetali.

L'introduzione di DNA esogeno all'interno di cellule di lievito prende il nome di **trasformazione**. Il processo è, infatti, essenzialmente analogo a quello applicato per la trasformazione batterica, benché la procedura sia diversa. Le cellule vengono trattate in modo tale da alterare la permeabilità sia della parete che della

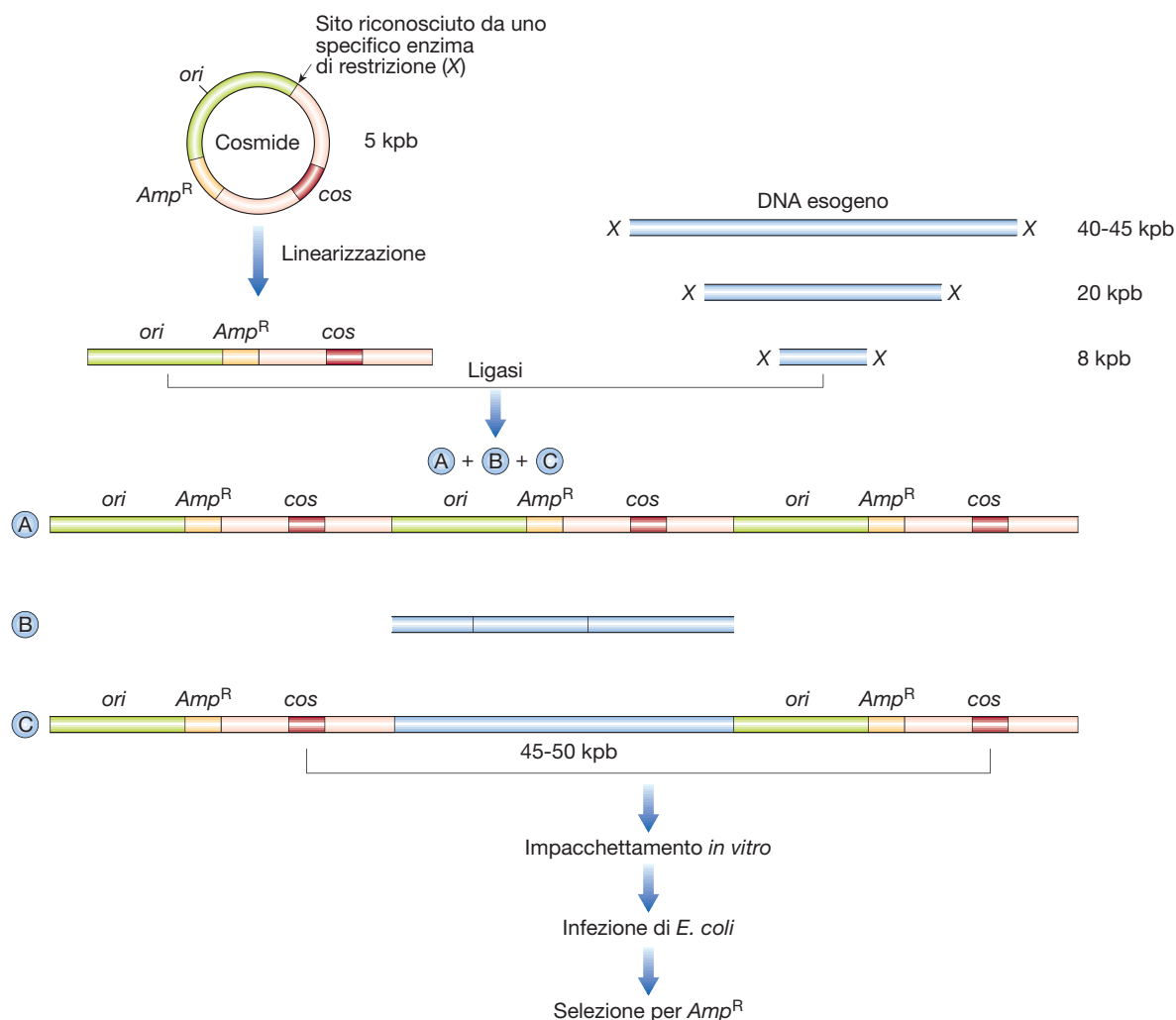


Figura 20.23 Schema della procedura di clonaggio in un cosmide.

In alto a sinistra è schematizzato un tipico cosmide di circa 5 kbp con i suoi elementi genetici caratterizzanti: *ori*, origine di replicazione; *Amp^R*, gene che conferisce la resistenza all'ampicillina; *cos*, sequenze coesive del fago λ ; con la freccia è indicato il sito riconosciuto da uno specifico enzima di restrizione (X). In alto a destra sono indicati tre ipotetici frammenti di DNA esogeno di diverse dimensioni generati con il taglio da parte dell'enzima di restrizione X. Dopo linearizzazione del vettore cosmidico mediante taglio con l'ER X e liguazione

in vitro con frammenti esogeni con estremità X si possono generare i prodotti indicati in A, B e C. I prodotti A e B corrispondono, rispettivamente, alla formazione di possibili oligomeri tra le molecole di cosmidi linearizzati o tra i frammenti di DNA esogeno. Il prodotto C corrisponde, invece, alla possibile formazione di molecole di DNA ricombinanti tra il vettore cosmidico e un frammento di DNA esogeno: il cosmide ricombinante può essere impacchettato *in vitro* se le dimensioni di DNA compreso tra due sequenze *cos* sono quelle previste (45-50 kbp) per entrare nella testa del batteriofago.

membrana, così da permettere l'introduzione di DNA esogeno senza influenzare, però, in modo drastico la vitalità delle cellule e permettendo la ricostituzione delle strutture alterate.

L'inserimento di DNA esogeno nelle cellule animali è, normalmente, un processo a due stadi. Ancora una volta, alterando la permeabilità delle cellule con alcuni "trucchi" sperimentali, il DNA esogeno viene trasferito all'interno delle cellule: nel caso di cellule animali tale trasferimento prende il nome di **trasfezione**, perché il termine "trasformazione" era già da anni usato in ambiente medico per indicare la trasformazione oncogenica di una cellula. Se il DNA esogeno trasferito all'interno di una cellula non è integrato all'interno del genoma della cellula trasfettata si parla di **trasfezione transiente**: la cellula manterrà il DNA in essa introdotto solo per un periodo di tempo limitato e verrà perso dopo poche divisioni cellulari. Durante tale periodo, però, se il DNA esogeno contiene un gene d'interesse sotto il controllo di

un opportuno promotore, la proteina codificata da quel gene può transitoriamente essere espressa nelle cellule trasfettate, permettendo, così, una serie di studi funzionali. È però anche possibile che il DNA esogeno sia integrato all'interno del genoma delle cellule trasfettate: questo rappresenta il secondo stadio indicato sopra e, in tal caso, si parla di **trasfezione stabile**. Nella maggior parte dei casi, l'integrazione del DNA esogeno all'interno del genoma avviene a caso, secondo meccanismi di **ricombinazione non omologa** o **illegittima**. Negli ultimi anni sono state sviluppate tecniche molto importanti che permettono di dirigere l'integrazione del DNA esogeno nel genoma in una posizione prestabilita: tali tecniche sono alla base della costruzione di **animali transgenici** in cui l'espressione di un gene è specificamente eliminata (*vedi* cap. 22).

Le tecniche con cui si riesce a introdurre DNA nelle cellule animali o vegetali sono le più svariate:

- introduzione diretta mediante meccanismi di endocitosi favoriti dalla precipitazione del DNA con sali di calcio;
- trasfezione mediata da vescicole lipidiche, chiamate liposomi, che contengono il DNA;
- permeabilizzazione delle cellule facilitata da shock elettrici (elettroporazione);
- microiniezione diretta del DNA nella cellula;
- metodi biolistici in cui il DNA viene fatto assorbire su microsfere che poi vengono letteralmente “sparate” all'interno delle cellule.

Infine, come vedremo appresso, in molti casi il DNA esogeno viene inserito all'interno di cellule eucariotiche animali o vegetali attraverso l'utilizzo di virus.

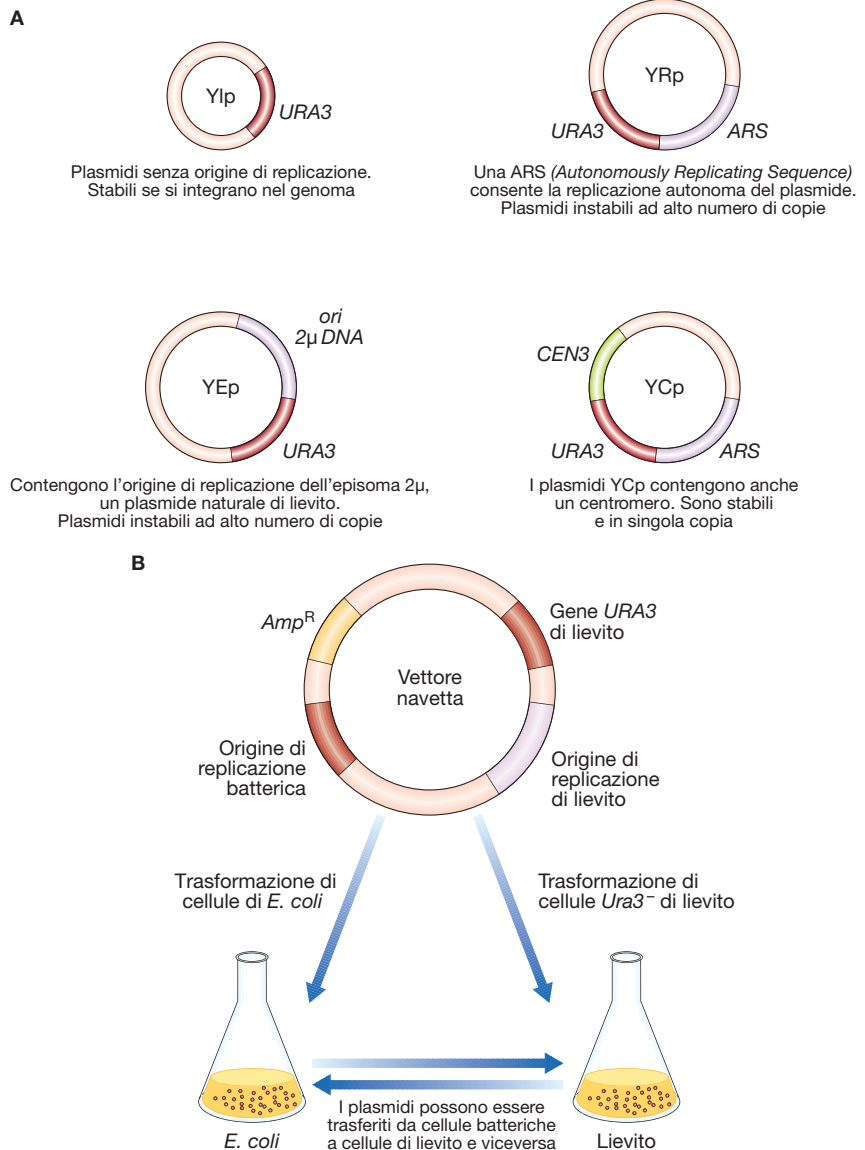
Vettori di lievito

Il lievito *Saccharomyces cerevisiae* è uno degli organismi modello più frequentemente utilizzati per studi di Biologia Molecolare negli eucarioti (*vedi* cap. 22). I vettori per il clonaggio e/o l'espressione di geni in lievito sono molto simili ai vettori plasmidici usati nei batteri; differiscono, però, per i marcatori utilizzati per selezionare i trasformanti, per la presenza o assenza di origini di replicazione e di altre sequenze cromosomiche, quali centromeri e telomeri, che influenzano profondamente la stabilità dei vettori e le dimensioni degli inserti di DNA che questi possono contenere (**fig. 20.24**).

I vettori per il clonaggio in lievito devono possedere alcune delle caratteristiche generali già precedentemente descritte per i vettori batterici, quale la presenza di siti di restrizione opportuni, magari condensati in un frammento MCS. Devono possedere, inoltre, un marcatore di selezione per individuare le cellule trasformate che hanno trasferito al loro interno il vettore di clonaggio o il vettore ricombinante contenente il gene d'interesse. Come marcatori di selezione in lievito raramente si utilizzano marcatori che conferiscono resistenza ad antibiotici o altre sostanze. Più spesso si utilizzano dei marcatori metabolici, quali i geni codificanti per enzimi richiesti per la sintesi di uno specifico amminoacido o di una base azotata (ad es. *LEU2*, richiesto per la sintesi di leucina, o *URA3*, richiesto per la sintesi di uracile). Ovviamente, come ceppi utilizzati per la trasformazione si utilizzeranno cellule di lievito mutate nei geni utilizzati come marcatori sul plasmide (nell'esempio citato sopra ceppi di lievito *Leu2⁻* o *Ura3⁻*, incapaci di crescere, rispettivamente, in terreni privi di leucina o di uracile). La presenza del plasmide che porta i geni *LEU2* o *URA3* selvatici permetterà alle cellule mutanti trasformate di crescere anche in assenza di leucina o uracile.

I classici vettori plasmidici per lievito si distinguono in quattro categorie: **YIp** (*Yeast Integrating plasmid*), **YRp** (*Yeast Replicative plasmid*), **YEpl** (*Yeast Episomal plasmid*) e **YCp** (*Yeast Centromeric plasmid*). Tutti e 4 i tipi di plasmide (**fig. 20.24A**) contengono un marcatore di selezione, ma i plasmidi YIp non hanno un'origine di replicazione, mentre gli altri 3 hanno un'origine di replicazione corrispondente a

Figura 20.24 Vettori plasmidici con diverse caratteristiche possono essere utilizzati per il clonaggio nel lievito *Saccharomyces cerevisiae*. (A) Sono indicati i 4 tipi di vettori plasmidici (YIp, YRp, YE_p, YCp) più frequentemente utilizzati per esperimenti di clonaggio nel lievito *S. cerevisiae*. Le loro caratteristiche principali sono indicate in figura e, più dettagliatamente, nel testo. (B) Caratteristiche di un vettore navetta che può essere trasferito da cellule batteriche a cellule di lievito e viceversa.



un'origine cromosomica (ARS nei plasmidi YRp e YCp) o all'origine di un episoma di lievito chiamato 2μ (YE_p).

I plasmidi YIp, essendo privi di un'origine di replicazione, non possono replicarsi autonomamente e, quindi, vengono per lo più persi dalla cellula nel corso delle divisioni cellulari, a meno che non vengano integrati da qualche parte in uno dei cromosomi di lievito. Nel lievito, il meccanismo d'integrazione avviene essenzialmente solo per ricombinazione omologa: ne deriva che i plasmidi YIp possono ricombinare e, quindi, integrarsi o in corrispondenza di regioni di omologia nel marcatore di selezione o in corrispondenza dell'inserito di DNA portato dal plasmide ricombinante, ammesso che tale inserito contenga sequenze di DNA di lievito. Proprio per tale motivo, il lievito è l'organismo d'elezione per esperimenti di knock-out genetico.

Gli altri tipi di plasmide contengono un'origine di replicazione e sono, quindi, in grado di replicarsi autonomamente una volta entrati per trasformazione all'interno di cellule di lievito. La loro stabilità e il loro numero di copie variano a seconda del tipo di origine di replicazione presente sul plasmide e, in particolare, i plasmidi del tipo YCp sono molto stabili e presenti in una sola copia perché

contengono la sequenza corrispondente a un centromero di lievito e, quindi, segregano in modo appropriato sia in mitosi che in meiosi.

Infine, ai vari tipi di vettori per lievito vengono quasi sempre aggiunti marcatori che possono essere selezionati in cellule di *E. coli* (ad es. *Amp^R*) e un'origine di replicazione funzionale nei batteri: tali vettori sono chiamati **shuttle vector** o **vettori navetta**, perché possono essere propagati indifferentemente sia in cellule batteriche che di lievito (**fig. 20.24B**).

I vettori descritti sopra permettono il clonaggio di frammenti di DNA da poche centinaia di basi fino a un massimo di 10-20 kpb. Sono stati sviluppati anche vettori di clonaggio per lievito molto particolari, chiamati **YAC** o **Yeast Artificial Chromosome**, che permettono il clonaggio di frammenti molto grandi di DNA genomico fino a 500-1000 kpb.

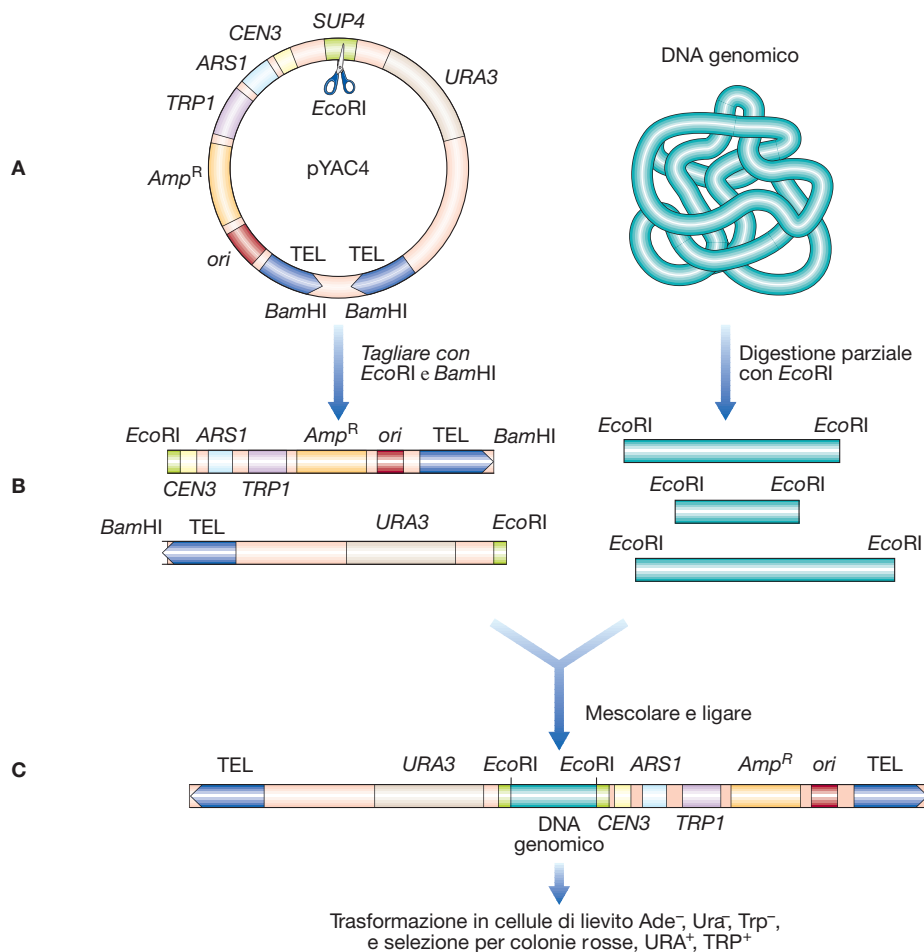


Figura 20.25 Clonaggio di DNA in un vettore YAC per lievito.

(A) Il taglio con gli enzimi di restrizione *EcoRI* e *BamHI* divide il DNA circolare del vettore YAC nelle due braccia mostrate in (B), ciascuna contenente un marcatore selettivo per lievito e almeno un elemento cromosomico (TEL). Il DNA da clonare è ottenuto tagliando il genoma tramite digestione parziale con *EcoRI*. L'inserimento di frammenti *EcoRI* nel sito *EcoRI* del vettore YAC porta all'inattivazione del gene per il tRNA soppressore *SUP4*. Se il tRNA soppressore è funzionante, viene inserito un residuo dell'amminoacido tirosina in corrispondenza di una mutazione non senso nel gene di lievito mutato *ade2* presente nel ceppo usato per la trasformazione. La mutazione nel gene *ade2* determina l'accumulo di un pigmento rosso, per cui le cellule *ade2* mutanti danno origine a colonie rosse: se il gene *SUP4* è presente nel vettore YAC, si ha la soppressione

dell'effetto della mutazione e le colonie avranno colore bianco; se *SUP4* è però inattivato dall'inserimento di un frammento *EcoRI* nei vettori YAC ricombinanti, le colonie saranno rosse. Le due braccia dello YAC contengono due marcatori per selezione diversi *TRP1* e *URA3*. Plasmidi YAC ricombinanti si generano soltanto se l'inserimento di frammenti *EcoRI* esogeni permette il ricongiungimento tra le due braccia del vettore. Gli YAC ricombinanti sono, quindi, trasferiti tramite trasformazione all'interno di cellule di lievito che portano mutazioni nei geni *TRP1*, *URA3* e *ADE2*. Dopo trasformazione, gli YAC ricombinanti sono selezionati in quanto in grado di crescere su terreno privo di triptofano e di uracile e perché daranno origine a colonie di colore rosso: ciò perché il gene *SUP4* è inattivato dalla procedura di clonaggio e, quindi, è incapace di sopprimere l'effetto della mutazione *ade2*.

Questi vettori (**fig. 20.25**) contengono diversi elementi cromosomici e marcatori di selezione rappresentati in figura e qui sotto elencati:

- una sequenza corrispondente a un centromero di lievito (ad es. CEN3);
- una sequenza corrispondente a un'origine di replicazione cromosomica (ad es. ARS1);
- sequenze corrispondenti ai telomeri di lievito (TEL);
- i geni *TRP1* e *URA3* che permettono a cellule *Trp1⁻* e *Ura3⁻* di crescere in assenza di triptofano o uracile;
- il gene *SUP4*, che, come descritto nella didascalia di **figura 20.25**, codifica per un gene tRNA soppressore;
- geni richiesti per la selezione e propagazione in cellule batteriche, quali *ori* (origine di replicazione) e *Amp^R* (marcatore di resistenza all'ampicillina).

La procedura di clonaggio in un vettore YAC e la selezione dei cloni ricombinanti sono schematizzate in **figura 20.25** e descritte più in dettaglio nella didascalia.

Vettori per il trasferimento genetico in cellule animali

Una descrizione esauriente di tutti i possibili tipi di vettori per il trasferimento genico in cellule animali e vegetali va ben oltre lo scopo di questo testo; pertanto, in questo paragrafo sono indicati solo alcuni dei vettori o dei metodi più utilizzati, lasciando a letture specialistiche una trattazione più ampia dell'argomento. Inoltre, la descrizione di vettori e tecniche utilizzate per la costruzione di animali transgenici è presentata nel capitolo 22.

La replicazione autonoma tipica dei plasmidi non funziona in cellule di mammifero. Alcuni vettori sono però in grado di effettuare una replicazione plasmide-simile se contengono l'origine di replicazione di un virus. In **figura 20.26** è mostrato un vettore episomale che è in grado di replicarsi in alcuni tipi di cellule di mammifero in coltura, dato che nel plasmide è inserita l'origine di replicazione del virus SV40 (*Simian Virus 40*). Il vettore contiene altri elementi genetici richiesti per l'espressione del gene clonato nel vettore e per la selezione delle cellule contenenti il DNA ricombinante. Spesso tali vettori contengono altri elementi genetici per la propagazione in cellule batteriche e, quindi, ricadono nella categoria dei vettori navetta già descritta nell'ambito dei vettori per cellule di lievito.

Ci sono altri tipi di vettori per cellule di mammifero, e molti di questi sono basati sull'utilizzo di vettori retrovirali (**fig. 20.27**). I vettori retrovirali derivano principalmente dall'oncoretrovirus MoLV (*Moloney Leukemia Virus*) e sono in grado di accogliere fino a circa 7 kpb di DNA esogeno contenente il transgene. Tali vettori retrovirali sono però in grado di trasfettare solo cellule in attiva proliferazione, limitandone l'utilizzo. Per ovviare a questo problema, sono stati costruiti vettori derivati da lentivirus, come HIV, che hanno sviluppato l'abilità di integrarsi anche nel genoma di cellule non proliferanti, grazie alla capacità del provirus di attraversare la membrana nucleare. I retrovirus e il loro ciclo replicativo è stato descritto nel capitolo 8 (*vedi* finestra 8.3). Varie modifiche di tali retrovirus hanno permesso di costruire vettori lentivirali contenenti meno del 10% del genoma virale iniziale in cui sono presenti solo le sequenze necessarie a retrotrascrivere, trasferire e integrare la cassetta di espressione nelle cellule bersaglio e sono assenti i geni per le proteine virali patogene. La costruzione di questi vettori ha raggiunto buoni livelli di biosicurezza, grazie anche all'inattivazione del promotore virale originale presente nelle LTR (*Long Terminal Repeat*). Uno svantaggio dei vettori derivati da virus a integrazione, come MoLV o HIV, è rappresentato dal fatto che l'integrazione nel genoma avviene in modo non controllato e ciò può comportare effetti di mutagenesi inserzionale. Può, infatti, accadere che il vettore si inserisca all'interno di un gene cellulare essenziale, determinandone la perdita di funzione e provocando la morte della singola cellula. Può anche accadere che il vettore entri all'interno di geni inattivandoli, oppure in prossimità di altri geni alterandone i profili di espressione in

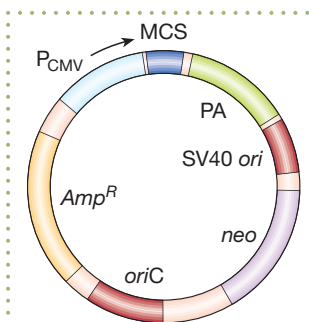


Figura 20.26 Schema di un vettore episomale per la trasfezione transiente di cellule di mammifero. Nel vettore sono presenti i seguenti elementi: MCS rappresenta il Multiple Cloning Site, contenente una serie di siti di restrizione per il clonaggio di inserti di DNA; il gene contenuto nell'inserto può venire espresso in quanto è posto sotto il controllo del promotore del citomegalovirus virus (*P_{CMV}*); PA indica un frammento che contiene il segnale di poliadrenilazione che conferisce stabilità al trascritto; *SV40 ori* è la sequenza di DNA corrispondente all'origine di replicazione del virus SV40; *neo* è il gene per la selezione delle cellule trasfettate che, codificando per la neomicina fosfotransferasi, conferisce alle cellule di mammifero la resistenza all'antibiotico G418; *oriC* indica l'origine di replicazione riconosciuta in cellule di *E. coli*; *Amp^R* è il gene che conferisce a cellule di *E. coli* la resistenza all'ampicillina.

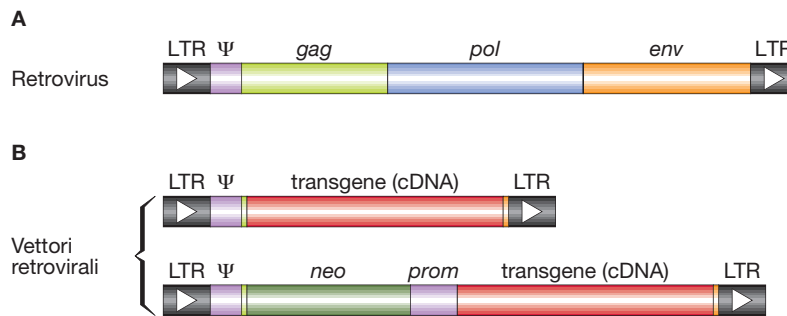


Figura 20.27 Vettori retrovirali. (A) Il genoma dei retrovirus, qui presentato nella forma di DNA a doppia elica integrata nel genoma ospite, contiene almeno i tre geni *gag*, *pol* ed *env*, le lunghe ripetizioni terminali (LTR) e la sequenza Ψ (psi) necessaria per l'impacchettamento del virione. (B) Nei vettori retrovirali i geni virali sono stati sostituiti dal transgene di interesse (in genere un

cDNA) che può essere trascritto a partire dal promotore/enhancer presente nella LTR di sinistra, oppure da un promotore (*prom*) appositamente posizionato a monte del transgene. In alcuni casi è presente anche un marcatore selezionabile, per esempio la resistenza alla neomicina (*neo*).

modo potenzialmente pericoloso per la cellula. Tutte queste problematiche vanno tenute in attenta considerazione nella scelta dell'utilizzo di vettori retrovirali per il trasferimento genico in cellule di mammifero.

Trasferimento di DNA nei vegetali

Il trasferimento di geni all'interno di organismi vegetali apre prospettive molto importanti per aumentarne, per esempio, le proprietà nutrizionali o la possibilità di favorirne la crescita in condizioni ambientali sfavorevoli. Un'importante proprietà delle piante riguarda il loro grado di plasticità, per cui una cellula vegetale è, in opportune condizioni, in grado di rigenerare una pianta completa. È pertanto possibile manipolare singole cellule per cercare di produrre piante con nuove proprietà.

Le piante non contengono plasmidi naturali, ma il **plasmide batterico Ti** del microrganismo del suolo *Agrobacterium tumefaciens* è utilizzato per introdurre geni esogeni in cellule vegetali (**fig. 20.28**). Senza entrare in dettagli, basti ricordare che una lesione sullo stelo di una pianta permette l'infezione da parte di *A. tumefaciens* e l'introduzione del plasmide Ti, una molecola molto grande (circa 200 kpb). Dopo l'infezione, parte del plasmide Ti chiamata **T-DNA**, si integra nel genoma della pianta in modo apparentemente casuale. Il T-DNA contiene geni, alcuni dei quali responsabili dell'insorgenza di "tumori" nella pianta. Al fine di utilizzarlo per introdurre geni esogeni il T-DNA è stato modificato così da perdere la capacità d'indurre tumori; infatti, il DNA esogeno può essere inserito tra le sequenze confine di destra e di sinistra indicate in **figura 20.28** che determinano l'integrazione nel genoma della pianta senza causare alcun tumore. Anche nel caso di cellule vegetali è necessario riconoscere le cellule che hanno introdotto plasmidi Ti modificati: si utilizzano, quindi, marcatori selettivi. Il marcatore solitamente più utilizzato conferisce la resistenza alla kanamicina, anche se marcatori che danno resistenze ad altri antibiotici o erbicidi sono altrettanto usati. Ancora una volta, perché il DNA esogeno possa essere espresso, è necessario che sia trascritto sotto il controllo di un promotore che funziona nelle piante.

Il sistema di *A. tumefaciens* è, tuttavia, applicabile solo a un numero limitato di specie vegetali. Metodi alternativi per l'introduzione di DNA esogeno prevedono la trasformazione di protoplasti, cioè di cellule vegetali private della parete cellulare, o l'introduzione di DNA con tecniche biolistiche tramite l'utilizzo di microsfele di tungsteno od oro ricoperte di DNA e "sparate" all'interno della cellula. Anche virus che infettano cellule vegetali sono utilizzati per introdurre DNA al loro interno. Uno tra i più utilizzati è il **virus del mosaico del cavolfiore (CaMV)**, il cui genoma è costituito da una molecola di DNA circolare a doppio filamento

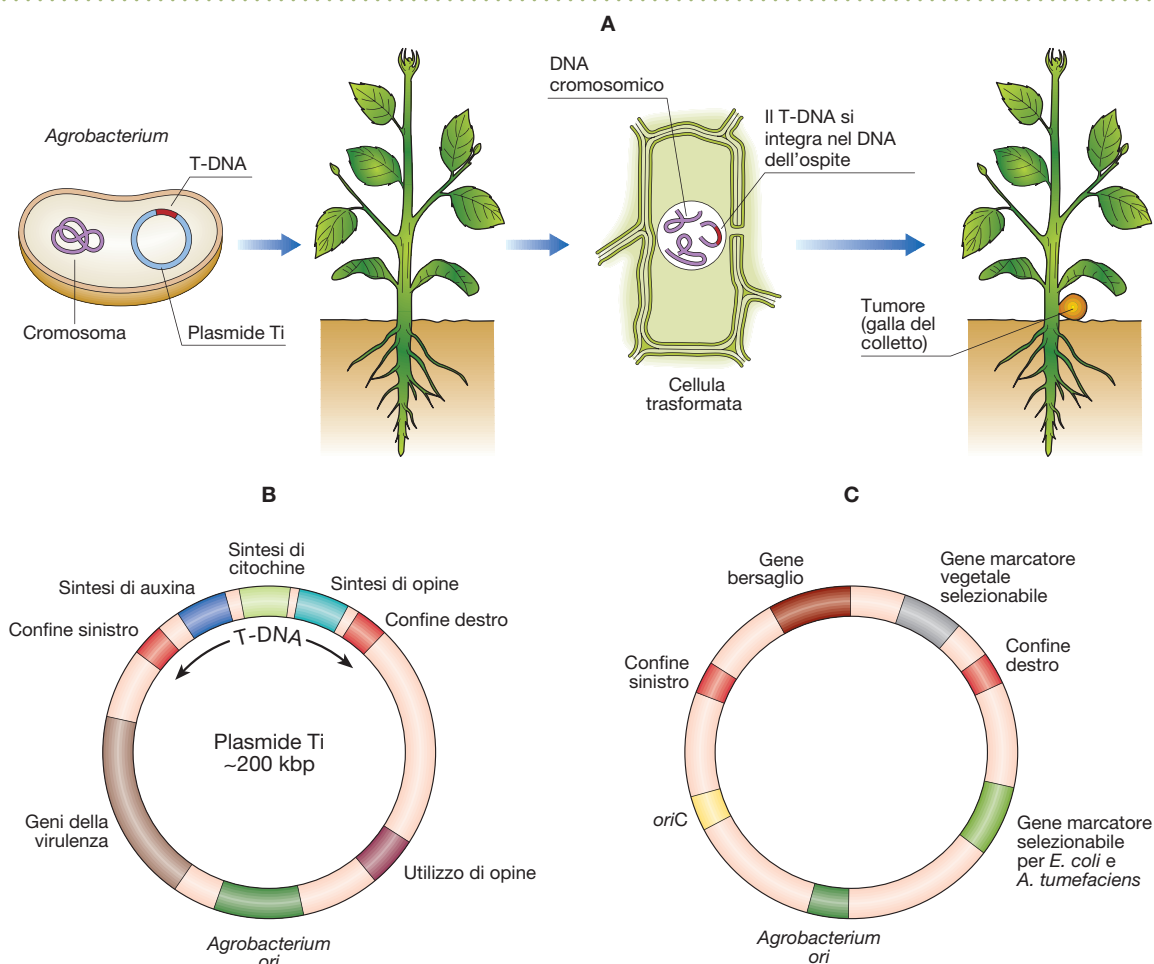


Figura 20.28 *Agrobacterium tumefaciens* e il suo plasmide Ti come vettore di clonaggio nelle cellule vegetali. (A) Una ferita sullo stelo di una pianta sana permette l'infezione da parte di *A. tumefaciens* e la formazione di tumori (galle) sullo stelo della pianta. (B)

Il plasmide naturale Ti presente in *A. tumefaciens* con le regioni geniche più rilevanti. (C) Un vettore plasmidico Ti ricombinante con gli elementi genetici più rilevanti.

contenuta all'interno di un involucro proteico che, come il DNA del batteriofago lambda, impone un limite alle dimensioni del DNA esogeno che può essere inserito nel genoma virale senza che questo perda la capacità di infettare le cellule.

20.5 Banche di DNA

Banche genomiche

Abbiamo talvolta utilizzato le espressioni “clonare un gene” o “esprimere un gene”: nel primo caso si intendeva dire che si era riusciti a clonare un frammento di DNA contenente un intero gene, nel secondo caso che si era riusciti addirittura a trascrivere il gene in esame e a fare produrre la proteina eventualmente codificata dal gene stesso.

Il problema principale però è: da che cosa si parte per identificare e isolare il gene d'interesse? Fino a pochi anni fa si partiva esclusivamente da collezioni di DNA chiamate **librerie**, **banche** o **genoteche**. Una libreria di DNA è, infatti, una collezione di frammenti di DNA clonati in uno specifico vettore di clonaggio nella quale l'insieme di tutti i cloni rappresenta l'intero genoma dell'organismo da cui il DNA è stato isolato. Tale tipo di libreria viene definita genomica per distinguerla

da un altro tipo di libreria costituita da copie di DNA (cDNA) delle molecole di mRNA presenti nelle cellule dell'organismo d'interesse. Vedremo più avanti come si costruiscono copie di DNA partendo da molecole di mRNA, e quali sono le differenze più importanti nell'utilizzo di librerie genomiche e di cDNA.

Riteniamo importante descrivere i metodi di costruzione e l'utilizzo di librerie di DNA per l'isolamento e il clonaggio di geni, perché tali approcci sono stati essenziali per comprendere a livello molecolare la funzione dei geni. Tuttavia, in questi ultimi anni, le librerie di DNA sono sempre meno utilizzate. Infatti, lo sviluppo della tecnica nota come *Polymerase Chain Reaction* o PCR (che descriveremo più avanti) permette di sintetizzare *in vitro* e clonare qualunque segmento di DNA la cui sequenza sia nota.

Per costruire una **libreria genomica**, il primo passo consiste nel decidere in quale vettore di clonaggio si vuole costruire la libreria. Abbiamo già detto che i diversi vettori hanno proprietà specifiche e, tra queste, una caratteristica importante riguarda le dimensioni dei frammenti di DNA che nel vettore possono essere clonati. Il secondo passo consiste nella frammentazione del DNA genomico nelle dimensioni prescelte per la costruzione della libreria. Il DNA si può frammentare casualmente, per esempio facendo passare più volte la preparazione di DNA attraverso l'ago di una siringa o trattando la preparazione di DNA con ultrasuoni. Più le condizioni utilizzate sono drastiche, più i frammenti di DNA saranno piccoli: la loro dimensione presenterà una distribuzione gaussiana la cui media può venire sperimentalmente determinata mediante elettroforesi su gel (**finestra 20.3**). È possibile frammentare il DNA mediante trattamento con enzimi di restrizione specifici: più la sequenza di DNA riconosciuta da un enzima è lunga e più i frammenti di DNA ottenuti saranno di grandi dimensioni. Va ricordato che gli enzimi di restrizione tagliano il DNA non in modo casuale, ma in corrispondenza di specifiche sequenze nucleotidiche: ciò ha notevoli implicazioni, tra cui, per esempio, il fatto che se un certo gene contiene al suo interno la sequenza riconosciuta da un dato ER, non si potrà generare un singolo frammento di DNA contenente l'intero gene.

Per la costruzione di banche di DNA e per il loro successivo utilizzo è importante generare una collezione di frammenti parzialmente sovrapposti: ciò permette, infatti, di allineare i frammenti ottenuti in modo contiguo uno all'altro (**fig. 20.29**). Per ottenere tale risultato spesso si utilizza il taglio con enzimi

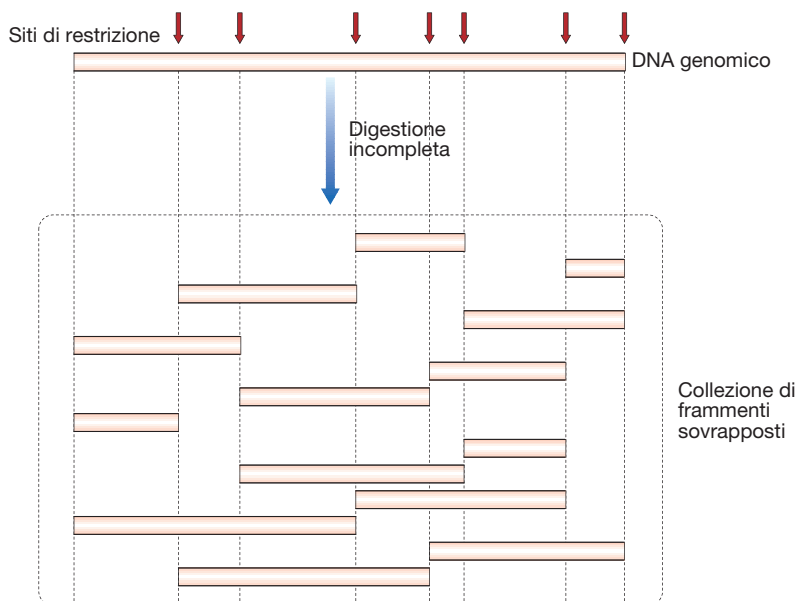


Figura 20.29 Preparazione di frammenti sovrapposti di DNA genomico mediante digestione parziale con un enzima di restrizione. La digestione parziale di DNA genomico con un enzima di restrizione che riconosce i siti indicati dalle frecce genera una collezione di frammenti sovrapposti, esemplificata dallo schema nel riquadro in basso. La digestione completa che si ottiene tagliando tutti i siti indicati dalle frecce rosse genererebbe, invece, frammenti di DNA non sovrapposti.

di restrizione, ma avendo l'accortezza di operare in modo tale da generare una **digestione parziale del DNA**. Ciò significa utilizzare condizioni sperimentali (tempi brevi di digestione o quantità limitanti di enzima) che causino il taglio solo di una frazione dei siti di restrizione presenti sul DNA (**fig. 20.29**). Il DNA digerito viene poi separato tramite elettroforesi per selezionare frammenti delle dimensioni medie prescelte per il clonaggio nel vettore preferito.

Rappresentatività di una libreria di DNA genomico

Un parametro importante da valutare per una libreria genomica è la sua rappresentatività, cioè la probabilità di trovarvi un clone contenente il gene d'interesse.

Facciamo un esempio: vogliamo trovare un certo gene all'interno di una libreria di *Drosophila melanogaster* e il genoma di tale organismo ha una dimensione di circa 2×10^8 kpb. Se frammentiamo il DNA di *D. melanogaster* in frammenti della dimensione media di 2×10^4 kpb, la libreria dovrebbe essere in grado di rappresentare l'intero genoma con 10 000 cloni indipendenti (2×10^8 kpb / 2×10^4 kpb). Ciò sarebbe vero, però, se tutti i cloni fossero diversi e tutti i frammenti non fossero sovrapposti uno con l'altro: entrambi i presupposti, però, non sono veri se si assume che i frammenti siano stati prodotti dalla digestione parziale con enzimi di restrizione o mediante metodi di rottura casuale del DNA. In una collezione di frammenti casuali, più grande e completa è la libreria e maggiore è la sua rappresentatività. Non esiste la certezza assoluta che la libreria contenga il frammento d'interesse; si può, però, definire la probabilità che tale frammento sia presente nella libreria in esame.

Il numero di cloni indipendenti (N) che devono essere analizzati per trovare nella libreria, con una certa probabilità, il frammento d'interesse può essere calcolato con la formula:

$$N = \frac{\ln(1-P)}{\ln(1-f)}$$

dove N è il numero di cloni indipendenti, P è una probabilità (ad es. 90% con $P = 0,9$, oppure 99% con $P = 0,99$) ed f è la frazione del genoma contenuta in un clone medio della libreria che si ottiene dividendo la media delle dimensioni dei frammenti con la dimensione totale del genoma (nell'esempio sopra citato $2 \times 10^4 / 2 \times 10^8$).

Il numero di cloni indipendenti si riferisce, per vettori plasmidici, al numero di batteri che sono stati trasformati in origine o, nel caso del batteriofago lambda, al numero di particelle fagiche prodotte nella reazione di impacchettamento *in vitro*.

Il numero di cloni indipendenti della libreria che si dovranno analizzare (**tab. 20.3**) dipende, ovviamente, dalle dimensioni medie dell'inserto (f), mantenendo fisso il valore della probabilità (ad es. 99%).

Tabella 20.3 Stima del numero di cloni (N) da analizzare nel caso di librerie genomiche di uomo in diversi vettori.

Organismo	Dimensione del genoma aploide	Tipo	Dimensione dell'inserto (f)	Probabilità (P)	Numero di cloni indipendenti della libreria
Uomo	3×10^9 basi	Plasmide	4 kpb	0,99	$3,5 \times 10^6$
		Lambda	20 kpb	0,99	$7,7 \times 10^5$
		Cosmide	40 kpb	0,99	$3,5 \times 10^5$
		YAC	600 kpb	0,99	$2,3 \times 10^4$

Librerie di cDNA

Una libreria genomica rappresenta tutto il DNA di un dato organismo sia che questo contenga geni codificanti per proteine, sia geni per RNA, sequenze regolative, esoni o introni, DNA ripetuto ecc. Tuttavia, in molti casi, si cercano le porzioni dei geni che vengono espresse in proteine che, soprattutto negli organismi multicellulari, rappresentano una frazione molto piccola dell'intero genoma.

Per fare questo si prepara la libreria a partire dagli mRNA estratti dalle cellule d'interesse. Ne deriva, per esempio, che nelle librerie di cDNA mancheranno le sequenze di DNA corrispondenti agli introni, che, non essendo codificanti, non sono rappresentate negli mRNA maturi. Analogamente, mancheranno sequenze importanti per la regolazione dell'espressione di un gene, quali i promotori o gli enhancer. Infine, va ricordato che, a differenza di una libreria genomica, una libreria di cDNA rappresenta solo quei geni che sono effettivamente trascritti ed espressi nel tessuto o nel tipo cellulare utilizzato per la preparazione degli mRNA. Quindi, mentre una banca di DNA genomico, per esempio di topo, è la stessa indipendentemente dal tessuto utilizzato per preparare il DNA, una banca di cDNA preparata partendo da mRNA estratti dal fegato di un topo sarà diversa da una banca di cDNA preparata dal cervello dello stesso organismo, perché diversi sono i geni espressi nei due tessuti. Infine, mentre i cloni di una banca di DNA genomico sono statisticamente rappresentati in modo simile (un'eccezione riguarda cloni contenenti sequenze di DNA ripetute più volte), i diversi cloni di cDNA possono essere più o meno rappresentati in relazione al livello di espressione (e quindi al numero di molecole di mRNA) dei diversi geni. In pratica, geni altamente espressi saranno più rappresentati nella banca di cDNA rispetto a geni che sono trascritti a livelli più bassi.

Poiché le molecole di mRNA non possono essere clonate direttamente, il primo passo nella costruzione di una libreria di cDNA consiste nel sintetizzare una copia di DNA (cDNA) dei diversi mRNA.

Senza entrare in eccessivi dettagli sperimentali, la procedura di base è descritta in **figura 20.30**. Il primo passo nella costruzione della libreria consiste nella purificazione degli mRNA dalle altre forme di RNA presenti nelle cellule in esame che sono quantitativamente molto più abbondanti. Negli eucarioti si sfrutta la presenza all'estremità 3'-OH di una coda di poli-A nella maggior parte delle molecole di mRNA mature (*vedi* finestra 14.2).

L'enzima in grado di copiare molecole di mRNA in DNA si chiama **trascrittasi inversa**; abbiamo già descritto la funzione biologica di questa proteina nella replicazione dei retrovirus (*vedi* cap. 8 e poco sopra). La trascrittasi inversa è in grado di utilizzare molecole di RNA come stampo per la sintesi di un filamento di DNA complementare alla sequenza dell'RNA. Come le classiche DNA polimerasi ha, però, bisogno di un innesco che provveda l'estremità 3'-OH a cui aggiungere via via i deossinucleotidi. Come innesco si utilizzano molecole di oligo-dT che possono appaiarsi alla coda di poli-A presente nelle molecole di mRNA. Alla fine della reazione della trascrittasi inversa si forma una molecola ibrida contenente un filamento di RNA e un filamento di DNA.

Il filamento di RNA viene degradato mediante il trattamento dell'ibrido in ambiente alcalino (pH circa 12) o utilizzando degli enzimi chiamati RNasi H che degradano il filamento di RNA in una molecola ibrida a doppio filamento RNA-DNA. Le molecole di DNA a singolo filamento così prodotte tendono a formare delle strutture secondarie a forcina (hairpin) ripiegandosi su se stesse. Ciò crea una regione di DNA a doppio filamento con un'estremità 3'-OH libera che la DNA polimerasi di *E. coli* è in grado di allungare determinando la sintesi di una molecola di DNA a doppio filamento completa. Tale molecola presenta però un'ansa a un'estremità che può essere risolta trattando con una nucleasi chiamata S1 che ha la proprietà di tagliare il DNA a singolo filamento e che, quindi, rimuove l'ansa stessa.

Figura 20.30 Sintesi di una molecola di cDNA a doppio filamento partendo da una molecola di mRNA maturo.

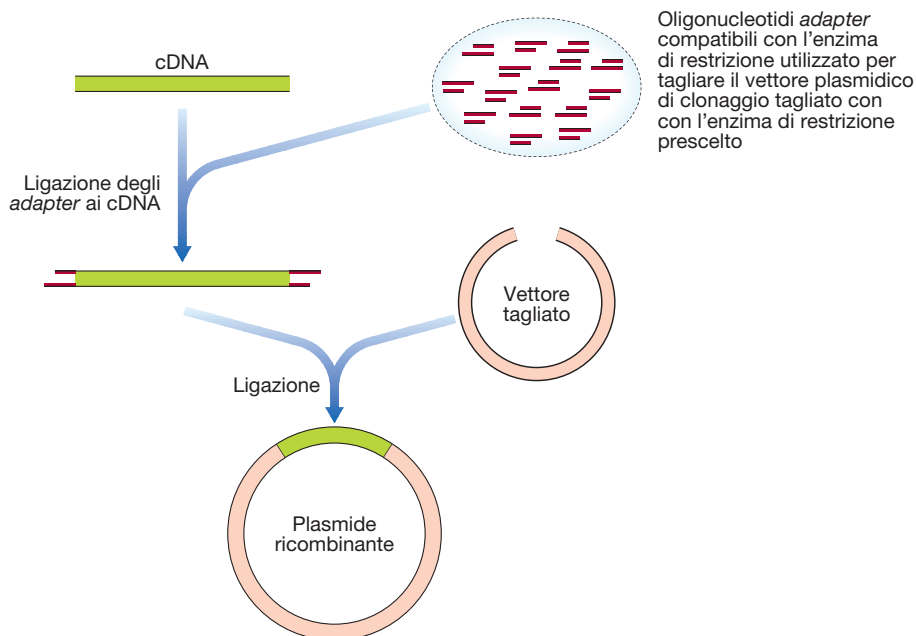
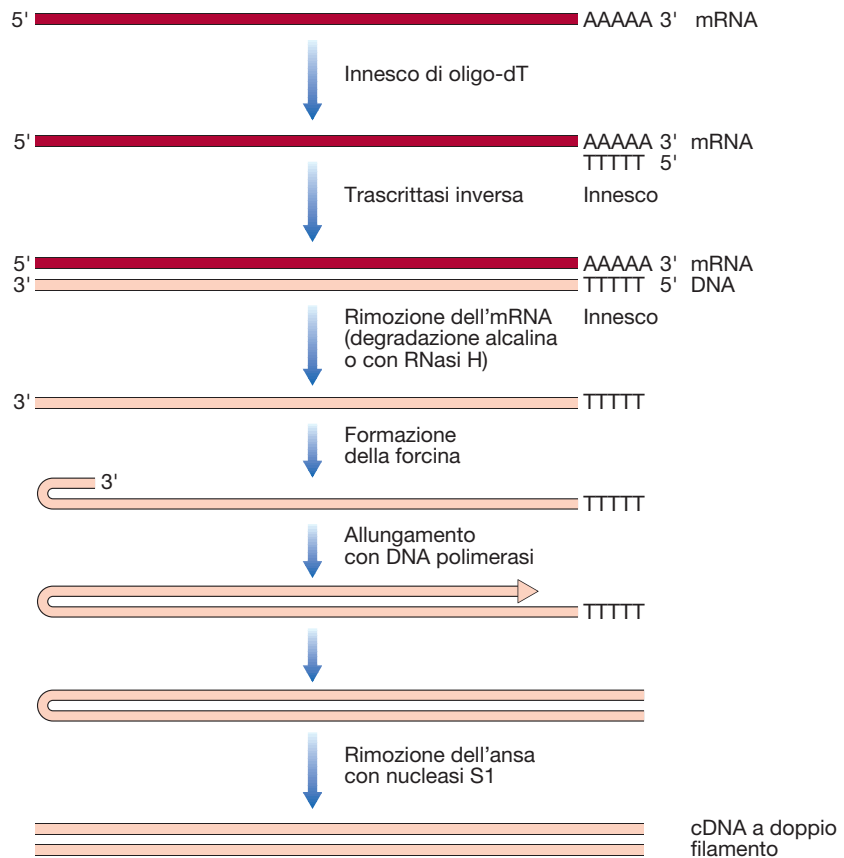


Figura 20.31 Clonaggio di cDNA in un vettore. È qui indicato lo schema di clonaggio di un cDNA in un vettore plasmidico. Alle estremità del cDNA vengono attaccati degli adattatori (*adapters*) o

giunti (*linkers*) sintetici che, dopo taglio con l'enzima di restrizione prescelto, possono generare delle estremità sporgenti compatibili per il clonaggio nel vettore plasmidico prescelto.

Le molecole di cDNA a doppio filamento così prodotte *in vitro* possono essere clonate in vettori di clonaggio attaccando alle loro estremità con la DNA ligasi corte molecole di DNA sintetizzate chimicamente e chiamate **adattatori** (*adapters*) o **giunti** (*linkers*). Questi oligonucleotidi sintetici hanno sequenze compatibili con i siti riconosciuti e tagliati da vari enzimi di restrizione e permettono, quindi, il clonaggio dei cDNA nel sito di restrizione prescelto per la costruzione della libreria di cDNA (**fig. 20.31**).

Identificazione del clone contenente il DNA d'interesse da una libreria

Una volta costruita una libreria genomica o di cDNA è necessario individuare tra le migliaia di cloni indipendenti che costituiscono la banca di DNA quale/i siano i cloni che contengono il gene d'interesse. In sostanza, si tratta di sviluppare rapidi metodi di "screening" tra moltissimi cloni.

Esistono una varietà di metodi di screening; ne descriveremo solo tre che sono molto diversi tra loro, ma che danno un'idea della diversità di approcci sperimentali possibili, lasciando ad altri testi specialistici una trattazione più ampia.

■ Screening funzionale mediante complementazione di uno specifico fenotipo

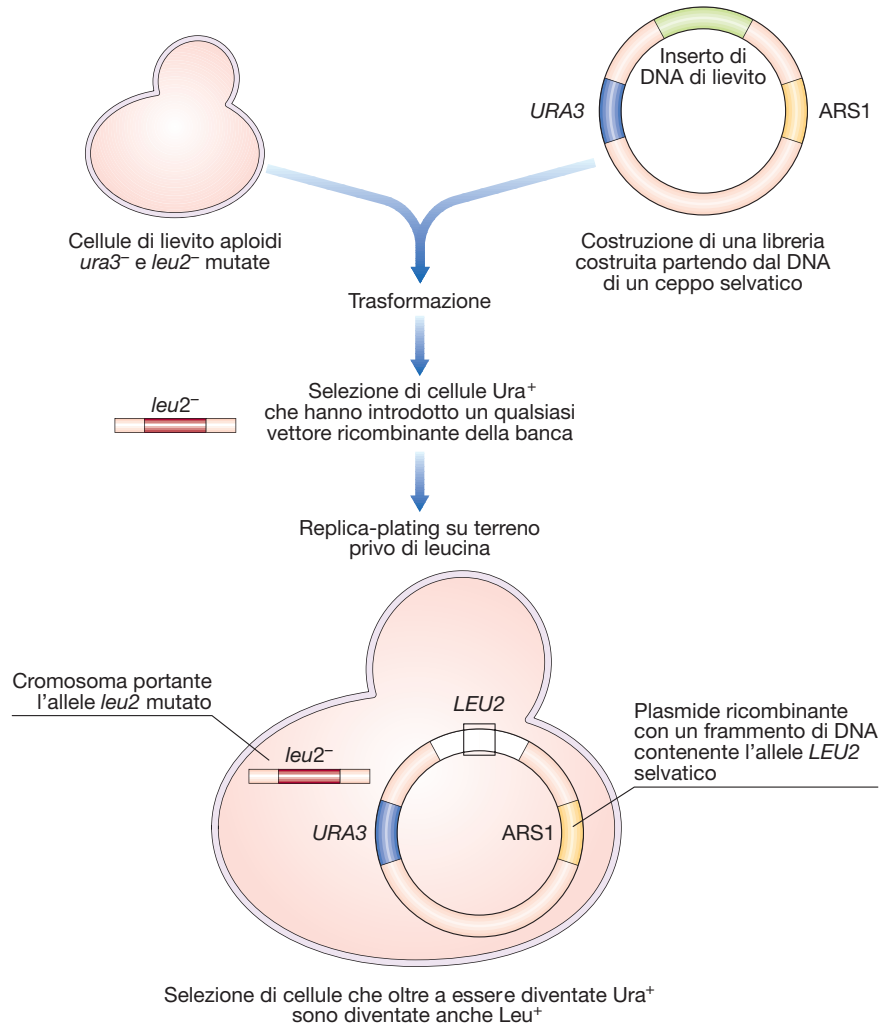
Tale approccio sperimentale si basa su concetti fondamentali della Genetica. Come abbiamo visto nel capitolo 7, la mutazione in un gene può causare uno specifico fenotipo: per esempio, una mutazione nel gene di lievito *LEU2* che codifica per un enzima richiesto per la biosintesi dell'amminoacido leucina può conferire alle cellule di lievito mutate un fenotipo *Leu⁻*: le cellule sono, cioè, incapaci di crescere in un terreno privo di leucina in quanto non sono in grado di produrre tale amminoacido, che deve, pertanto, essere addizionato al terreno per permettere la crescita cellulare. Analogamente, una mutazione, per esempio, nel gene *RAD2* che codifica per un enzima richiesto per riparare lesioni causate dal trattamento con raggi ultravioletti (UV), può rendere le cellule mutate molto più sensibili al trattamento con raggi UV rispetto alle cellule selvatiche. La maggior parte delle mutazioni sono recessive, ma poiché le cellule di lievito possono essere mantenute sia allo stato aploide che diploide (*vedi cap. 22*), una mutazione anche recessiva determina un fenotipo osservabile.

Come possiamo clonare il gene selvatico se abbiamo a disposizione un ceppo di lievito mutante? Supponiamo di avere, come descritto sopra, delle cellule mutanti *leu2* aploidi incapaci di crescere su un terreno privo di leucina. Si costruisce una banca genomica di DNA dal ceppo selvatico in un vettore per lievito capace di replicarsi autonomamente in tale organismo e contenente, per esempio, il marcatore selettivo *URA3*: nella banca, dovrebbero essere presenti uno o più cloni contenenti il gene *LEU2* selvatico. Si trasformano, quindi, cellule di lievito *leu2 ura3* mutate con la banca di DNA: si selezionano i trasformanti che hanno introdotto un qualunque clone ricombinante perché saranno in grado di crescere formando colonie su terreno privo di uracile (tutti i plasmidi ricombinanti della banca contengono il marcatore selettivo *URA3*). Le colonie vengono poi replicate su terreno privo sia di uracile che di leucina: cresceranno solo le cellule che contengono nel vettore ricombinante della banca, oltre al marcatore selettivo *URA3*, anche un frammento di DNA contenente il gene *LEU2*. Infatti, il gene *LEU2* presente nell'inserto del vettore ricombinante è in grado di complementare il difetto determinato dalla mutazione (**fig. 20.32**).

Un procedimento del tutto analogo può essere utilizzato per clonare il gene *RAD2*: in questo caso, però, quello che si andrà a cercare è la capacità del gene selvatico presente nella libreria di complementare la sensibilità a radiazione UV delle cellule mutate.

Con opportune modificazioni, tali screening di complementazione funzionale sono applicabili a molti tipi di cellule: il pre-requisito di tale approccio sperimen-

Figura 20.32 Clonaggio di un gene selvatico mediante complementazione funzionale del fenotipo causato da una mutazione recessiva nello stesso gene. L'esempio presentato riguarda la complementazione funzionale della mutazione recessiva *leu2* nel lievito *Saccharomyces cerevisiae* da parte di un plasmide ricombinante, presente in una libreria di DNA preparata dall'organismo selvatico, che porta il gene *LEU2* selvatico.



tale è la disponibilità di mutazioni che causino uno specifico fenotipo che può essere complementato dall'allele selvatico.

■ Screening con sonde di acidi nucleici

Il principio alla base di questa tecnica è un'evoluzione del Southern blot e prevede che abbiate ben compreso i concetti di sonda e di ibridazione descritti precedentemente in questo capitolo. Il termine "Southern blot" deriva dal nome dello scienziato Edwin Southern che per primo lo applicò (**fig. 20.33**). Abbiamo già visto che frammenti di DNA possono essere separati tramite elettroforesi su gel in base alle loro dimensioni. Se si prepara, però, del DNA da cellule di mammifero e lo si taglia con un enzima di restrizione, si generano frammenti di diverse dimensioni così numerosi che non sono osservabili come frammenti distinti dopo elettroforesi. Come si può identificare, allora, un frammento che contiene una specifica sequenza di DNA?

I frammenti di DNA separati mediante elettroforesi sono trasferiti (blotting) mediante capillarità o attraverso l'azione di un campo elettrico su un filtro di nitrocellulosa o nylon su cui il DNA rimane attaccato e può essere denaturato *in situ* dopo opportuni trattamenti (ad es. bagnando il filtro con soluzioni a pH alcalino) (**fig. 20.33**). Il filtro, a differenza del gel usato per l'elettroforesi, può essere facilmente maneggiato e, dopo denaturazione del DNA, può essere im-

merso in una soluzione contenente molecole di DNA a sequenza nota che sono state marcate con isotopi radioattivi. Queste molecole di DNA radioattive sono utilizzate come “sonda” (**finestra 20.1**) per individuare frammenti di DNA che contengono la stessa sequenza di DNA presente nella sonda radioattiva.

Ciò si realizza mediante esperimenti di ibridazione DNA-DNA. Sia i frammenti di DNA da analizzare che la sonda di DNA radioattivo possono essere denaturati, cioè si possono separare i due filamenti di DNA complementari. Come abbiamo già visto, ciò si ottiene o alzando la temperatura o variando il pH in modo da distruggere i legami chimici che tengono i due filamenti appaiati tra loro. A questo stadio del processo, il filtro su cui sono immobilizzati i frammenti di DNA denaturati viene immerso in una soluzione contenente la sonda di DNA radioattiva, anch'essa denaturata. La denaturazione del DNA è un processo reversibile che viene ulteriormente favorito in appropriate condizioni sperimentali. Utilizzando tali condizioni, la sonda di DNA denaturata può rinaturarsi (si dice anche “ibridare”) con la sequenza di DNA complementare eventualmente presente su un frammento immobilizzato sul filtro. Dopo un tempo opportuno per permettere la rinaturazione e una serie di lavaggi per eliminare la sonda radioattiva non ibridata ad alcuna sequenza presente sul filtro, quest'ultimo viene posto in contatto con una normale lastra per raggi X. Gli elettroni emessi dalla sonda

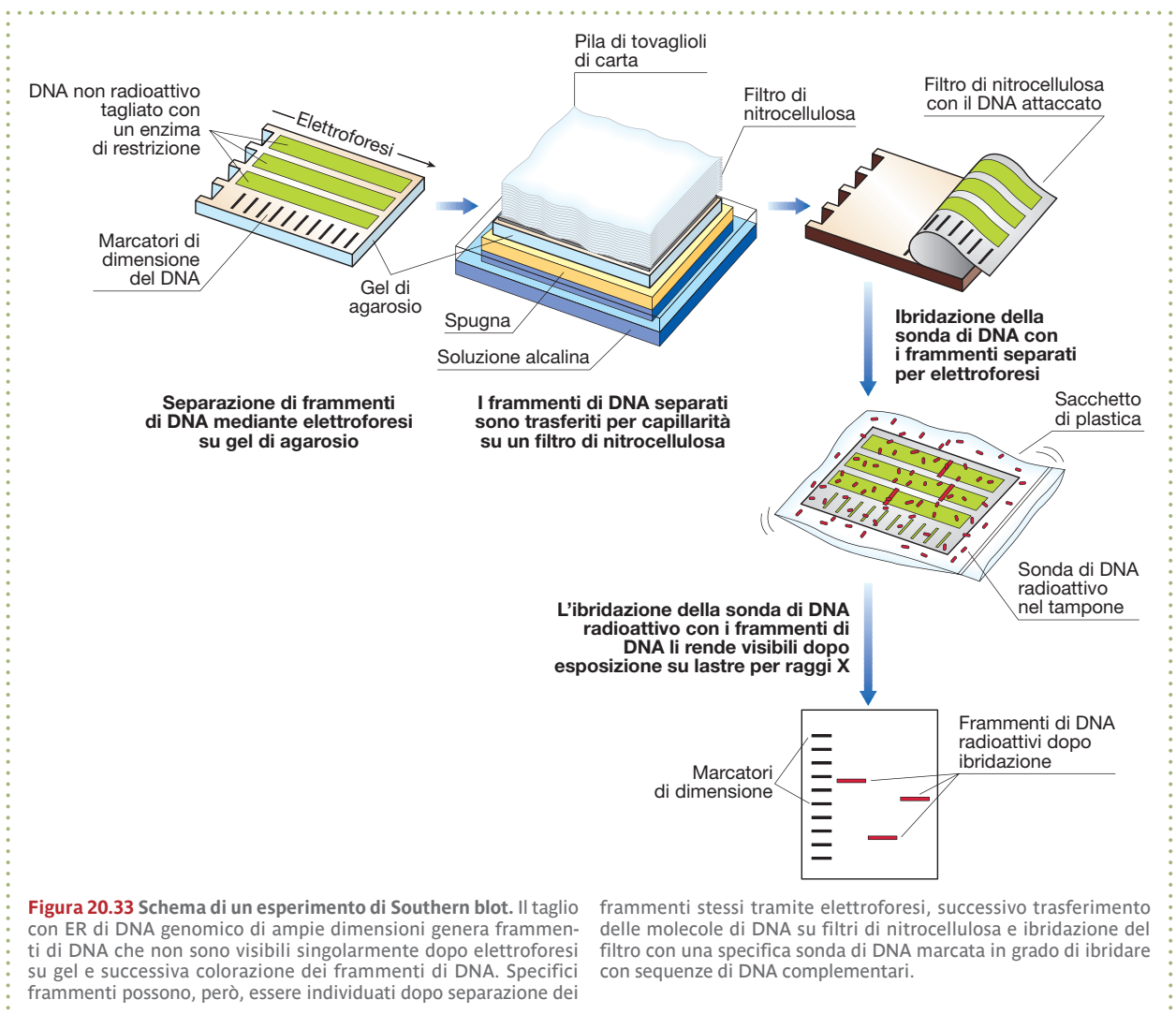


Figura 20.33 Schema di un esperimento di Southern blot. Il taglio con ER di DNA genomico di ampie dimensioni genera frammenti di DNA che non sono visibili singolarmente dopo elettroforesi su gel e successiva colorazione dei frammenti di DNA. Specifici frammenti possono, però, essere individuati dopo separazione dei

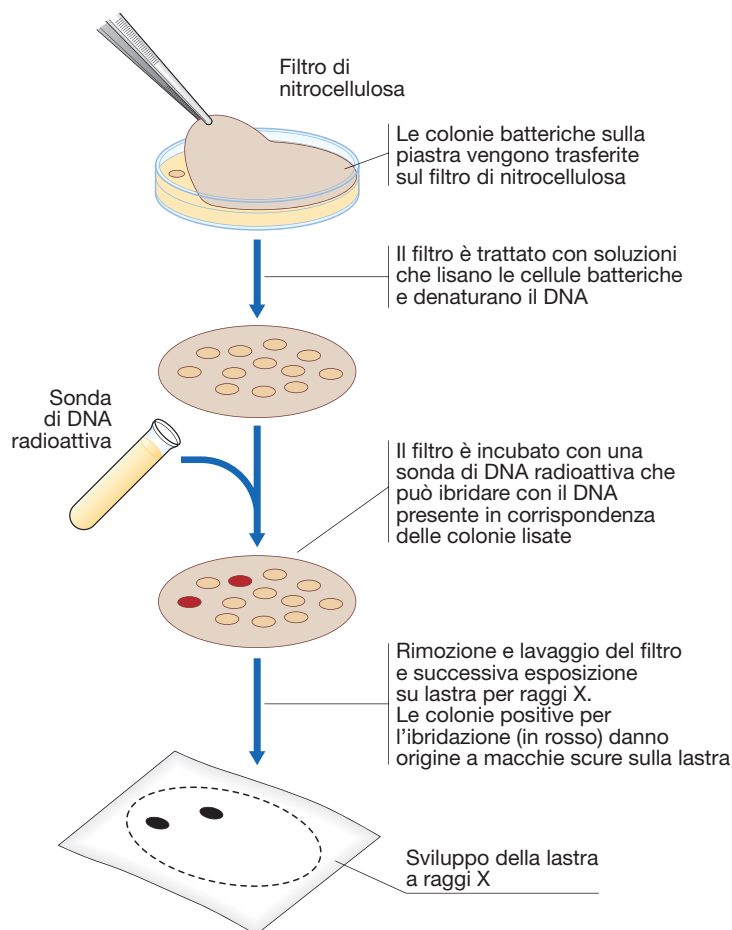
frammenti stessi tramite elettroforesi, successivo trasferimento delle molecole di DNA su filtri di nitrocellulosa e ibridazione del filtro con una specifica sonda di DNA marcata in grado di ibridare con sequenze di DNA complementari.

radioattiva ibridata sul filtro impressionano la lastra e il frammento di DNA ibridato dà così origine a un segnale che corrisponde a una banda scura (**fig. 20.33**).

La tecnica del **Southern blot** permette la visualizzazione di molecole di DNA mediante ibridazione DNA-DNA. Esistono altre tecniche di “blotting”, Northern e Western Blot, utilizzate per la visualizzazione, rispettivamente, di molecole di RNA e di proteine (**finestra 20.4**).

Lo screening dei cloni positivi di una libreria di DNA che identificano uno specifico gene segue la stessa logica del Southern blot. La tecnica sperimentale mostrata in **figura 20.34** prende il nome di **ibridazione su colonia**. In pratica, cellule di *E. coli* sono trasformate con la libreria di DNA preparata, nel caso mostrato, in un vettore plasmidico. Le cellule batteriche trasformate sono piastrate su un terreno selettivo e danno origine a delle colonie. Nella figura è mostrata una sola piastra con qualche colonia, ma nella pratica di laboratorio lo screening della banca prevede l'analisi di molti trasformanti per poter identificare il gene in esame: si dovranno, quindi, utilizzare molte piastre e su ciascuna di esse si può analizzare fino a qualche centinaia di colonie. Su ogni piastra è poi steso un filtro di nitrocellulosa sul quale parte delle cellule che formano ciascuna colonia rimane attaccata. I filtri sono, successivamente, stesi su della carta assorbente imbevuta con una soluzione che causa la lisi delle cellule batteriche depositate sul filtro e la denaturazione del DNA. I filtri corrispondenti alle varie piastre sono, poi, incubati con una soluzione contenente la sonda di DNA radioattiva corrispondente a parte del gene in esame. La sonda può ibridare con il DNA denaturato presente sul filtro nelle posizioni corrispondenti alle colonie originarie. I filtri, infine, sono rimossi, lavati

Figura 20.34 Clonaggio di un gene attraverso “ibridazione su colonia” di una libreria di DNA con una sonda di DNA radioattiva, corrispondente a un tratto del gene in esame. Un numero adeguato di cellule di *E. coli* trasformate con la libreria di DNA in cui, con una certa probabilità, sarà presente il gene d'interesse, è piastrato su capsule Petri con terreno solido. Dopo formazione delle colonie, sulle piastre sono stesi dei filtri di nitrocellulosa a cui si attacca parte delle cellule delle colonie; i filtri sono numerati e orientati in modo da poter risalire alla posizione delle colonie originarie. Dopo lisi delle cellule trasferite sui filtri e denaturazione del DNA, i filtri sono ibridati con una sonda di DNA radioattiva corrispondente a un tratto del gene che si vuole individuare nella libreria. La sonda potrà ibridare con le colonie che contengono plasmidi ricombinanti con inserti di DNA che presentano regioni di omologia rispetto alla sonda utilizzata. I segnali d'ibridazione positivi vengono rilevati su lastre per raggi X.



per eliminare la sonda residua non ibridata sul filtro e stesi su lastre per raggi X. Sulle lastre si formeranno delle macchie scure, dove la sonda radioattiva ha ibridato con il DNA corrispondente alle colonie che contenevano il vettore ricombinante della banca con inserito il gene d'interesse. Poiché la posizione delle macchie sulla lastra è riconducibile, tramite opportune numerazioni e un corretto orientamento nello spazio, alla posizione delle colonie sulle piastre, è possibile risalire alle colonie originali presenti sulle piastre madri in cui la libreria era stata distribuita. Le cellule di tali colonie vengono fatte crescere e da esse si isola il plasmide ricombinante: tali plasmidi contengono inserti di DNA in cui è incluso il gene d'interesse che è stato in grado di ibridare con la sonda radioattiva.

■ Screening di una libreria in un vettore di espressione mediante l'utilizzo di anticorpi

Come accennato precedentemente, è possibile utilizzare vettori d'espressione che portano alla produzione di proteine intere o proteine di fusione. Il vettore fagico noto come **λgt11** è tra i vettori più utilizzati per la costruzione di librerie di cDNA che possono portare alla produzione di proteine di fusione. Una descrizione della struttura di λgt11 e del suo utilizzo per la costruzione di librerie d'espressione e l'identificazione di parte del gene d'interesse utilizzando anticorpi prodotti contro la proteina codificata da quel gene sono schematizzati in **figura 20.35**.

Il vettore λgt11 contiene al suo interno parte del gene (*lacZ*) codificante per la β-galattosidasi (β-gal) di *E. coli*, la cui espressione è sotto il controllo del promotore *lac*. All'interno del gene codificante per la β-gal c'è un sito di riconoscimento per l'enzima di restrizione *EcoRI*. È possibile costruire una libreria di cDNA in λgt11 tagliando il vettore con *EcoRI* e inserendo mediante ligazione in tale sito le molecole di cDNA preparate dall'organismo prescelto a cui sono stati attaccati alle estremità dei linker *EcoRI* per permettere l'operazione di clonaggio. Se un dato cDNA si inserisce nel vettore mantenendo il modulo di lettura richiesto per la produzione corretta della proteina da esso codificata (1 probabilità su 6), inducendo la trascrizione dal promotore *lac*, si formerà dopo trascrizione e traduzione una proteina di fusione costituita, nella parte ammino-terminale, da una porzione di β-gal e, nella parte carbossi-terminale, da una porzione della proteina codificata dal cDNA d'interesse.

Come si possono riconoscere i fagi ricombinanti che danno origine alla proteina di fusione cercata? Dopo impacchettamento *in vitro* dei fagi, la libreria viene piastrata su una serie di capsule Petri ricoperte da uno strato di cellule batteriche sensibili al batteriofago lambda. Dopo infezione delle cellule di *E. coli*, ciascuno di essi darà origine a una placca di lisi visibile come una zona chiara sullo strato di batteri, a causa della lisi dei batteri infettati. A questo punto si procede in modo analogo alla procedura di ibridazione su colonia descritta poco sopra, eccetto che, operando su placche di lisi di fagi, si parla di **ibridazione su placca** e l'individuazione dei fagi ricombinanti che producono la proteina di fusione si basa sull'utilizzo di specifici anticorpi prodotti contro la proteina codificata dal gene d'interesse.

Filtri di nitrocellulosa sono stesi sulle diverse piastre e, dopo rimozione dei filtri, analogamente a quanto accade per il DNA, le proteine presenti in corrispondenza delle placche di lisi rimangono attaccate ai filtri. I filtri sono a questo punto incubati con anticorpi prodotti contro la proteina d'interesse. Se il cDNA codificante per la proteina si era inserito nel vettore in modo da originare una proteina di fusione con la sequenza amminoacidica corretta, tale proteina di fusione sarà riconosciuta dagli anticorpi. Si formano così sui filtri, in corrispondenza di una o più placche, dei complessi antigene-anticorpo che possono essere visualizzati con un anticorpo secondario marcato (fluorescente o radioattivo) che riconosca l'anticorpo specifico che si è legato alla proteina di fusione. Se, per esempio, gli anticorpi contro la proteina di fusione sono stati prodotti immuniz-

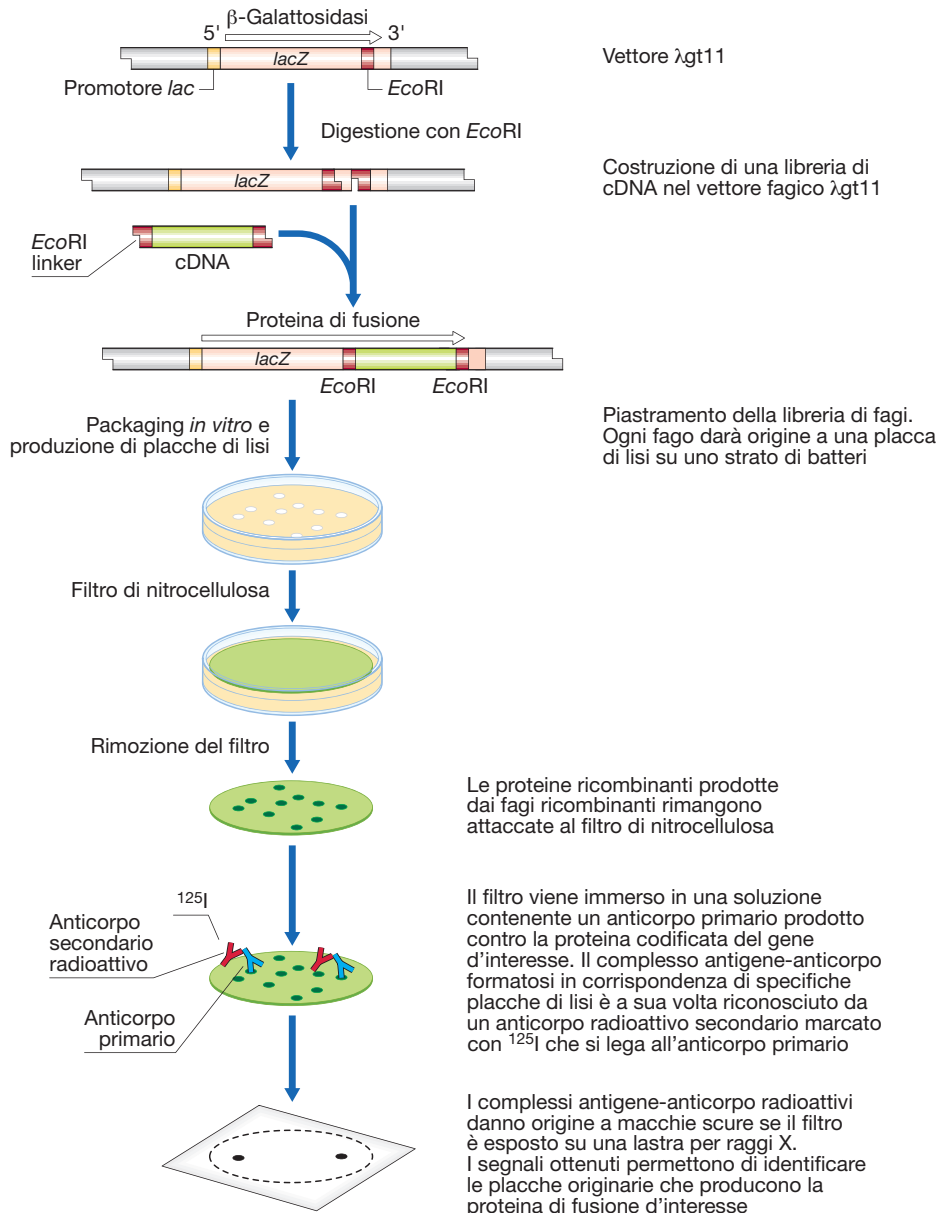


Figura 20.35 Utilizzo di una libreria nel vettore d'espressione λ gt11 per il clonaggio del gene codificante per una proteina contro la quale siano disponibili degli anticorpi. La disponibilità di anticorpi contro una proteina permette il clonaggio del gene codificante per la proteina stessa. Il clonaggio del gene richiede la costruzione di una libreria di cDNA in un opportuno vettore d'espressione: il vettore λ gt11 è tra quelli più utilizzati a tale scopo.

Come indicato in figura e nel testo, la costruzione della libreria e la sua espressione portano alla formazione di proteine di fusione nelle placche che esprimono cloni ricombinanti della libreria. La disponibilità di anticorpi contro la proteina d'interesse e la successiva visualizzazione dei complessi antigene-anticorpo formati permettono d'individuare le placche che esprimono la proteina di fusione e il successivo clonaggio di almeno parte del gene codificante.

zando dei conigli, gli anticorpi secondari marcati possono essere anticorpi prodotti in capra contro immunoglobuline di coniglio. Proseguendo nell'esempio, si è quindi formato in corrispondenza di alcune placche un complesso ternario costituito da: proteina di fusione/anticorpo primario (coniglio)/anticorpo secondario radioattivo (capra) e la radioattività rilasciata da quest'ultimo sarà in grado di impressionare le lastre per raggi X che sono state poste sui filtri. Ancora una volta sarà possibile, dai segnali lasciati sulla lastra, risalire alla posizione delle placche positive originarie e all'analisi del DNA in esse presente.

Va sottolineato che, con tale procedura, si identifica solo una parte del gene d'interesse; tuttavia, questo segmento del gene può, ora, essere utilizzato come sonda di DNA per individuare il gene completo, per esempio attraverso lo screening di una libreria genomica con la tecnica della ibridazione su colonia descritta precedentemente.

Questa breve panoramica rappresenta solo un "assaggio" delle diverse strategie messe in atto dai ricercatori per identificare il gene d'interesse da una libreria e dimostra come le tecnologie del DNA ricombinante siano tra loro diversificate e adattabili ai diversi reagenti e strumenti a disposizione dei ricercatori.

20.6 PCR: reazione a catena della polimerasi

La tecnica della PCR (*Polymerase Chain Reaction*, reazione a catena della polimerasi) è stata ideata nel 1983 da Kary B. Mullis e ha presto assunto un'importanza straordinaria per le moltissime applicazioni non solo in campo biologico o medico. Questa tecnica permette di selezionare e amplificare, da una preparazione eterogenea di DNA, un particolare tratto che ci interessa, producendone un grandissimo numero di copie. In sostanza, si tratta di una replicazione *in vitro* del frammento di DNA ottenuta mediante una DNA polimerasi e due oligonucleotidi lunghi 20-30 nt che fungono da inneschi (primer) per la replicazione e che sono responsabili della specificità della amplificazione. È importante notare che per poter fare una PCR è necessario conoscere la sequenza del tratto di DNA da amplificare o, almeno, le sequenze a cui fare associare gli inneschi ai due estremi di esso.

Un ciclo di PCR consiste di tre passaggi:

1. denaturazione del DNA;
2. associazione (*annealing*) degli inneschi ai filamenti di DNA denaturato;
3. sintesi da parte della polimerasi.

Questo ciclo viene ripetuto numerose (25-40) volte. La **figura 20.36** illustra schematicamente i primi quattro cicli di una PCR. Il DNA di partenza viene denaturato e poi ibridato con i due inneschi che vanno a legarsi ciascuno su un filamento a un estremo della regione da amplificare: si noti le direzionalità 5' → 3' dei filamenti del DNA di partenza e degli inneschi. Utilizzando i nucleosidi trifosfati che sono aggiunti nella miscela, la DNA polimerasi allunga gli inneschi in direzione 5' → 3'. I primi tentativi di PCR vennero fatti utilizzando la DNA polimerasi I di *E. coli*, che però, come la maggior parte delle proteine, veniva inattivata alla temperatura di denaturazione del DNA. Ciò comportava che a ogni ciclo di amplificazione si doveva riestrarre il DNA e ripreparare la miscela di amplificazione aggiungendo DNA polimerasi fresca. La grande innovazione introdotta da Mullis, e che gli ha fruttato l'assegnazione del premio Nobel per la chimica nel 1993, è stata l'uso nel protocollo della PCR di DNA polimerasi estratte da organismi termofili: quella più comunemente usata è la Taq DNA polimerasi preparata dal batterio *Thermophilus aquaticus*. Per ovvie ragioni evolutive questo enzima è molto stabile e non viene inattivato alle temperature vicino a 100 °C utilizzate nella fase di denaturazione del DNA. In questo modo è possibile preparare una soluzione contenente tutti i componenti necessari (DNA di partenza, Taq polimerasi, inneschi, nucleosidi trifosfati) e procedere a numerosi cicli di amplificazione semplicemente variando la temperatura della miscela di reazione. Tutto questo viene fatto in appositi strumenti che sono in sostanza dei piccoli termostati in cui è possibile programmare in maniera opportuna innalzamenti e abbassamenti della temperatura.

Un esempio di ciclo di amplificazione è il seguente:

- **denaturazione:** 30 sec a 95 °C per denaturare il DNA;
- **associazione** (*annealing*): 30 sec a 40-60 °C per favorire l'associazione specifica degli inneschi (la temperatura deve essere di circa 5 °C al di sotto della T_m degli inneschi);
- **sintesi:** 2 min a 72 °C, che è la temperatura ottimale per la Taq polimerasi.

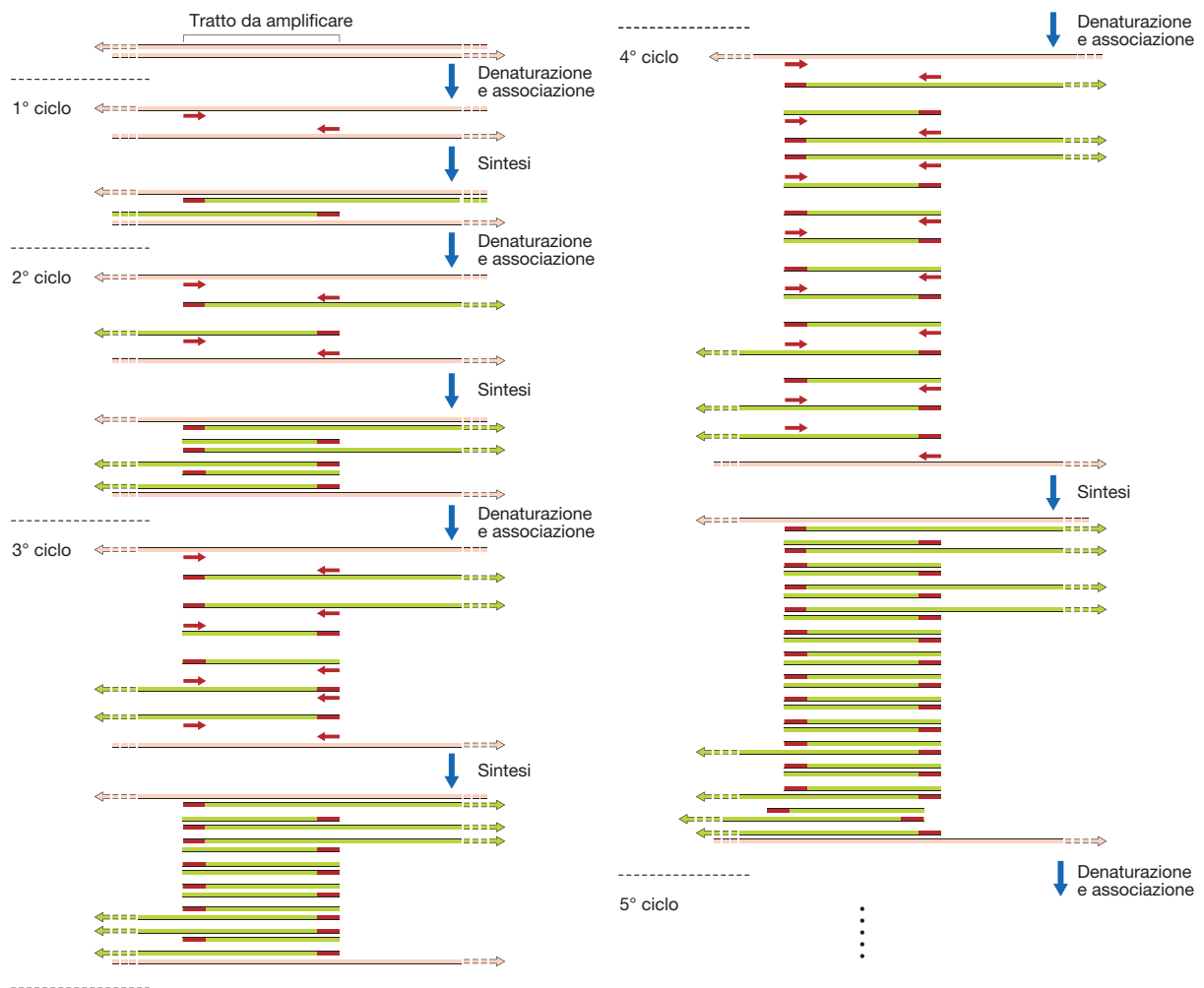


Figura 20.36 PCR: i primi quattro cicli. I filamenti di DNA di partenza sono indicati in rosa, gli inneschi (primer) in rosso e i filamenti prodotti durante l'amplificazione in verde. Dove i filamenti

di DNA e gli inneschi sono rappresentati come frecce, queste indicano la direzionalità 5'→3'.

I parametri (temperature, tempi e concentrazioni) di queste tre reazioni possono essere variati dal ricercatore per tenere conto di vari fattori, tra cui la quantità di materiale di partenza, la presenza di sequenze simili a quella di interesse, la composizione in basi degli inneschi, la lunghezza del tratto da amplificare ecc. Se si osserva con attenzione la **figura 20.36**, ci si può rendere conto che il DNA eterogeneo di partenza resta costante durante la PCR, le molecole che iniziano da uno degli inneschi e proseguono oltre l'altro innesco (amplificati lunghi) aumentano in maniera aritmetica (60 dopo 30 cicli) e le molecole amplificate tra i due inneschi (amplificati corti) aumentano in maniera esponenziale ($2^{30} \approx 10^9$, dopo 30 cicli).

Una variante della PCR che viene utilizzata per studiare l'espressione genica è la cosiddetta **RT-PCR**. Questa consiste di due passaggi: partendo da una preparazione di mRNA viene dapprima sintetizzato, utilizzando trascrittasi inversa (*Reverse Transcriptase* = RT), il cDNA, come abbiamo già descritto; successivamente, il cDNA viene utilizzato come materiale di partenza per una normale PCR usando inneschi specifici per la sequenza che desideriamo amplificare.

Queste tecniche, pur essendo molto efficienti, rapide e relativamente semplici, presentano alcuni problemi e limiti. Uno di questi è che, proprio a causa

dell'altissima efficienza di amplificazione che permette di partire da pochissimo materiale, sono facilmente soggette al rischio di contaminazione. Bastano poche molecole di DNA contaminante per ritrovarsi nel DNA amplificato sequenze non desiderate. Questi problemi sono ovviamente particolarmente gravi quando si parte da DNA danneggiato e/o da pochissimo DNA (ad es. DNA antico) che richiedono un alto numero di cicli di PCR.

Un'altra limitazione della normale tecnica di PCR è che essa è essenzialmente qualitativa: permette, cioè, di amplificare una sequenza di DNA se questa è presente nel campione in esame, ma non è in grado di darci una misura della quantità di tale sequenza (e quindi del DNA o RNA che la contiene), nel campione stesso. Infatti, col procedere dei cicli, l'amplificazione aumenta esponenzialmente fino a un certo punto per poi rallentare e, infine, fermarsi per esaurimento dei precursori trifosfati e/o degli inneschi. Il livello massimo raggiunto non è, quindi, proporzionale al numero di molecole di partenza. Poiché, invece, si vuole spesso misurare proprio la rappresentatività delle varie sequenze, soprattutto nell'analisi degli mRNA negli studi di espressione genica, sono state messe a punto delle variazioni sperimentali per rendere la PCR quantitativa.

Il metodo più interessante è chiamato **real-time PCR** (PCR in tempo reale). Con la real-time PCR, anziché andare ad analizzare il prodotto amplificato alla fine dei cicli, si segue il corso dell'amplificazione, ottenendo un profilo che descrive l'andamento della reazione stessa. Come si fa a seguire l'amplificazione, mentre essa avviene? Uno dei metodi escogitati consiste nell'aggiungere alla soluzione un composto fluorescente, per esempio è molto usato il *SYBR green*, che lega aspecificamente il DNA a doppio ma non a singolo filamento e quando è legato emette fluorescenza molto intensamente. Seguendo quindi spettroscopicamente la fluorescenza della soluzione dove avviene l'amplificazione, si otterrà una curva, come quelle mostrate in **figura 20.37**, che descrive quantitativamente l'andamento dell'amplificazione. Durante i primi cicli la quantità di DNA è ancora così bassa da non poter essere apprezzata; successivamente la curva inizia a salire con andamento esponenziale, raddoppiando a ogni ciclo. Dopo questa fase esponenziale l'aumento di DNA rallenta fino a fermarsi a un livello costante quando inneschi e/o nucleosidi trifosfati vengono esauriti. Se, come mostrato in **figura 20.37**, analizziamo diverse sequenze presenti nella stessa preparazione, troveremo che le curve di amplificazione, che alla fine della reazione raggiungono tutte più o meno lo stesso livello, differiscono per il momento in cui iniziano a salire. Infatti, una sequenza molto rappresentata nella preparazione iniziale di DNA comparirà dopo pochi cicli di amplifica-

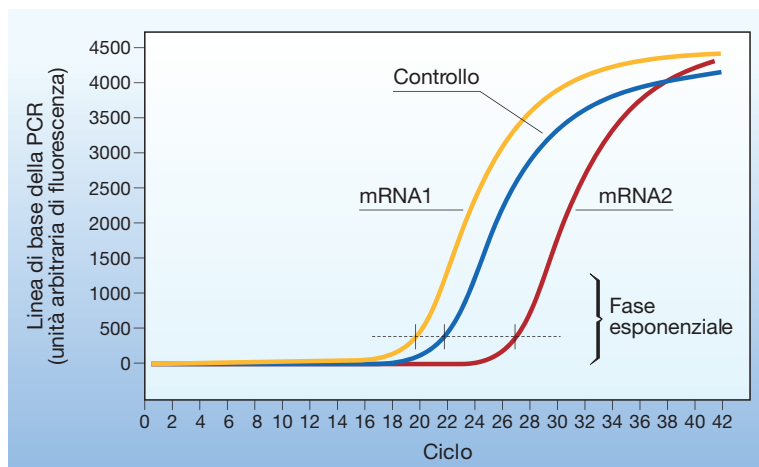


Figura 20.37 Real-time PCR. Profili di amplificazione da PCR delle sequenze corrispondenti a tre mRNA presenti nella stessa preparazione di partenza: mRNA1, mRNA2 e mRNA di controllo (si tratta quindi in realtà di una RT-PCR). La curva di amplificazione della sequenza corrispondente all'mRNA1 (giallo) sale 2 cicli di amplificazione prima della curva per l'mRNA di controllo (blu), indicando che ha una rappresentatività 4 (2^2) volte superiore al controllo; la curva relativa alla sequenza corrispondente all'mRNA2 (rosso) sale 5 cicli dopo la curva di controllo, indicando che ha una rappresentatività di circa 32 (2^5) volte inferiore al controllo.

zione, mentre una sequenza poco rappresentata comparirà più tardi. Come si vede nella figura, il confronto viene fatto prendendo in considerazione la prima parte delle curve, dove esse presentano un andamento esponenziale. Va sottolineato che la real-time PCR non permette di determinare il numero assoluto di copie di una particolare sequenza (DNA o RNA), ma permette di misurare il “numero relativo” di copie di una sequenza rispetto a un'altra. Infatti, viene sempre usata in confronti: per esempio, numero di copie di una sequenza in diversi tessuti, o diversi organismi, o in cellule in diverse condizioni, oppure differenti sequenze relativamente a una sequenza di controllo, come nell'esempio riportato in figura.

Le tecniche di PCR e RT-PCR e le numerose varianti che sono state escogitate sono oggi diventate strumenti di ricerca essenziali. Esse hanno, tra l'altro, in gran parte soppiantato le procedure classiche di clonaggio di DNA e cDNA. Queste, come abbiamo descritto precedentemente, richiedono due passaggi, ambedue lunghi, complessi e dispendiosi:

- clonaggio della preparazione eterogenea di partenza di DNA o cDNA;
- selezione del clone di interesse.

Le tecniche di PCR e RT-PCR permettono di ottenere la stessa cosa in un unico passaggio relativamente semplice. Infatti, la selezione viene fatta automaticamente durante l'amplificazione dalla specificità dei due inneschi che utilizziamo.

20.7 Determinazione della sequenza nucleotidica del DNA

A metà degli anni settanta sono stati escogitati due diversi metodi per “sequenziare” il DNA, cioè determinarne la sequenza delle basi (o nucleotidi). Il principio di base del sequenziamento, uguale per i due metodi, è illustrato in **figura 20.38**, dove, per semplicità, viene preso in considerazione un frammento di DNA a singolo filamento di 40 nucleotidi di lunghezza (in realtà, con opportuni accorgimenti, si possono analizzare frammenti a singolo o a doppio filamento lunghi centinaia di nucleotidi). Per determinare la posizione delle G (colorate di verde) in questo frammento, si inizia col generare, vedremo dopo come, l'insieme dei frammenti che si estendono dall'estremità 5' del filamento fino a ciascuna delle G interne (**figura 20.38A**). I frammenti tronchi ottenuti vengono poi separati con un sistema di elettroforesi ad alto voltaggio (in genere 2000 V) in un gel di poliacrilammide che sia in grado di separare frammenti che differiscono in lunghezza di un singolo nucleotide. La posizione delle bande nel gel ci permette di determinare le posizioni delle G nella sequenza di partenza (**fig. 20.38B**). Naturalmente, per avere la sequenza completa, la stessa procedura viene applicata, separatamente, per le quattro basi che sono poi corse in corsie adiacenti del gel, come mostrato nella **figura 20.38C**.

Come vengono generati i frammenti tronchi che si estendono da un punto fisso fino alle basi interne, specificamente per ciascuna base? Questo viene fatto in maniera molto diversa nei due metodi di sequenziamento.

Il **metodo** cosiddetto **chimico**, ideato da Walter Gilbert e Allan Maxam, si basa sul taglio del frammento di DNA (a singolo filamento) mediante specifiche reazioni chimiche, che qui non descriveremo in dettaglio. Queste permettono di modificare specificamente una delle quattro basi e, successivamente, di tagliare il filamento là dove si trova la base modificata. Questa frammentazione non viene fatta in maniera esaustiva, ma in maniera parziale in modo che ciascuna molecola subisca mediamente un singolo taglio in corrispondenza di una delle G (o di una delle altre basi in esame). I frammenti tronchi così ottenuti sono molti e non per-

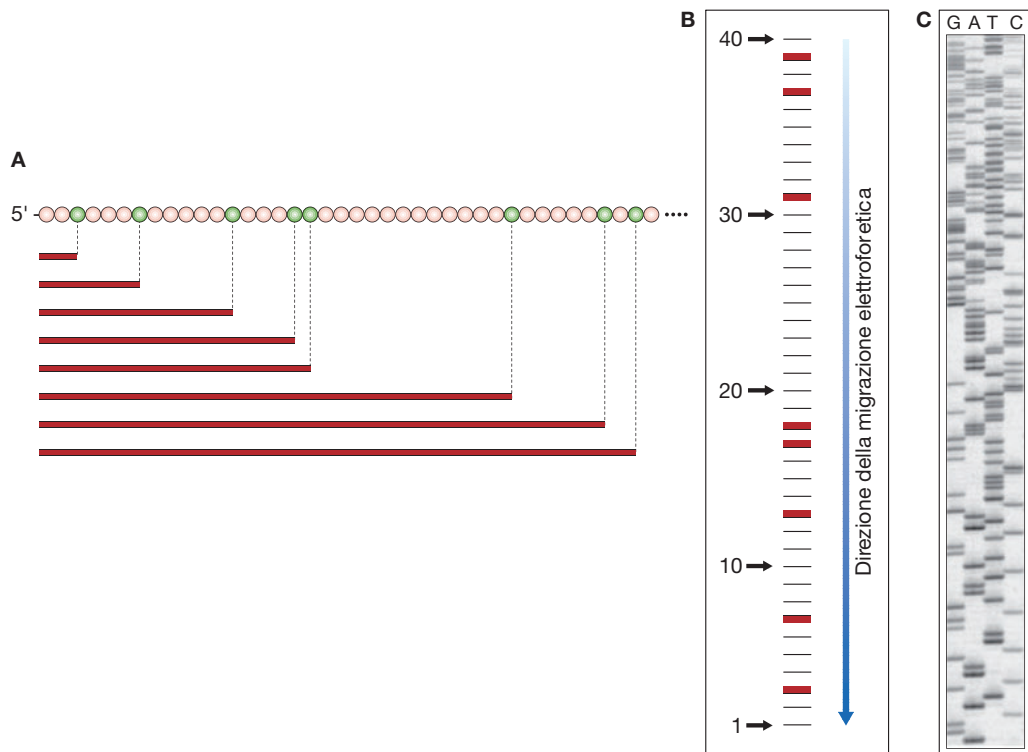


Figura 20.38 Principio su cui sono basati i due metodi di sequenziamento del DNA. (A) Consideriamo un frammento di DNA a singolo filamento, lungo 40 nt. Per determinare le posizioni in cui si trovano le G (verdi), si produce un insieme di frammenti tronchi che, partendo dall'estremità 5', si estendono fino a una delle G. (B) I frammenti tronchi vengono poi separati mediante elettroforesi ad alto voltaggio in gel di poliacrilammide, che è in grado di risolvere polinucleotidi le cui lunghezze differiscono anche di un

solo nucleotide. I frammenti più corti migrano più velocemente e si trovano quindi nella parte bassa del gel. Contando i nucleotidi dal basso verso l'alto possiamo determinare che le G si trovano nelle posizioni 3, 7, 13, 17, 18, 31, 37 e 39 a partire dall'estremità 5'. Naturalmente un'analoga procedura viene eseguita per le altre tre basi, A, T e C. (C) Come effettivamente si presenta un'autoradiografia di un gel di sequenza: dal basso verso l'alto (dal 5' verso il 3') si può leggere la sequenza: AGGCATGCAAGCTTGGC ecc.

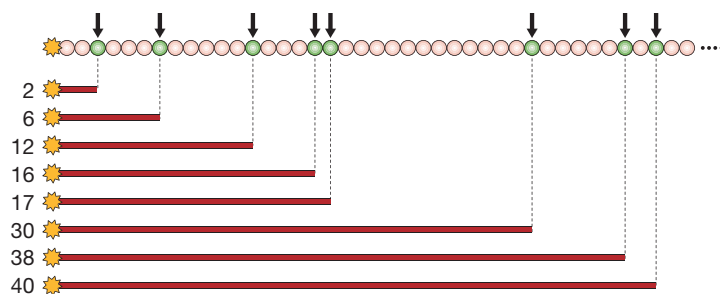


Figura 20.39 Produzione dei frammenti tronchi per il sequenziamento del DNA mediante il metodo chimico di Maxam e Gilbert. Il DNA a singolo filamento da sequenziare viene sottoposto a una reazione chimica che risulta nel taglio in corrispondenza di un solo tipo di base (in questo caso le G, cerchi verdi). La reazione è portata avanti in maniera parziale in modo tale che, statisticamente,

ogni molecola viene tagliata in corrispondenza di una sola G. Poiché il frammento da sequenziare è stato precedentemente marcato all'estremità 5' con un atomo di fosforo radioattivo (asterisco arancione), di tutti i frammenti tronchi che si formano solo quelli che partono dall'estremità 5' e si estendono fino a una delle G interne saranno visualizzati mediante autoradiografia del gel di sequenza.

metterebbero di decifrare la sequenza. Però se, come mostrato in **figura 20.39**, il filamento (singolo) da sequenziare viene prima marcato all'estremità 5' con un atomo di fosforo radioattivo, dopo l'elettroforesi vedremo, mediante autoradiografia, solo i frammenti tronchi radioattivi, cioè quelli che vanno dall'estremità 5' marcata radioattivamente fino a una delle G interne, individuando così tutte

le posizioni lungo il filamento dove si trovano le G. La stessa procedura viene fatta in provette separate anche per le A, le T e le C. Questo metodo è stato inizialmente preferito perché risulta in letture accurate (con pochi errori) e, inoltre, permette di sequenziare un filamento di DNA di cui non si sa nulla.

Il secondo **metodo**, chiamato **enzimatico** e ideato da Frederick Sanger, implica invece la sintesi dei frammenti tronchi da parte di una DNA polimerasi che utilizza come stampo il DNA di cui si vuole determinare la sequenza. Il frammento di DNA che si vuole sequenziare è prima clonato in un vettore: solitamente, all'interno di un *Multiple Cloning Site* o MCS di cui è nota la sequenza, in un vettore a singolo filamento, come quello derivato dal fago M13, perché è più facile sequenziare DNA a singolo filamento. È noto che le DNA polimerasi hanno, però, bisogno di un innesco per iniziare la sintesi. Come mostrato in **figura 20.40A**, l'innesco che introduciamo nella miscela di reazione sarà un oligonucleotide complementare alla sequenza già nota e adiacente alla regione da sequenziare. Come si fa in questo caso a produrre frammenti tronchi che si interrompano in corrispondenza di una base specifica, G, C, A, o T? Si aggiunge alla miscela di reazione, oltre ai normali quattro deossinucleosidi trifosfato, anche una piccola quantità (ad es. 1/500) di un particolare nucleotide il cui zucchero, anziché un normale deossiribosio, è un dideossiribosio. Come mostrato in **figura 20.40B**, questo manca del 3'-OH necessario per l'allungamento della catena. Quando durante la sintesi viene casualmente incorporato (una volta su 500) un dideossinucleotide, la catena si interrompe. In questo modo vengono sintetizzati tutti i possibili frammenti tronchi che iniziano al 5' con l'innesco e terminano su un nucleotide interno, G nel caso dell'esempio mostrato in figura. Naturalmente questo viene fatto, separatamente, per ciascuno dei quattro nucleotidi G, A, T o C. Nel metodo originale i frammenti tronchi prodotti erano radioattivi (o per la presenza nella miscela di reazione di un nucleoside trifosfato radioattivo o perché l'innesco utilizzato

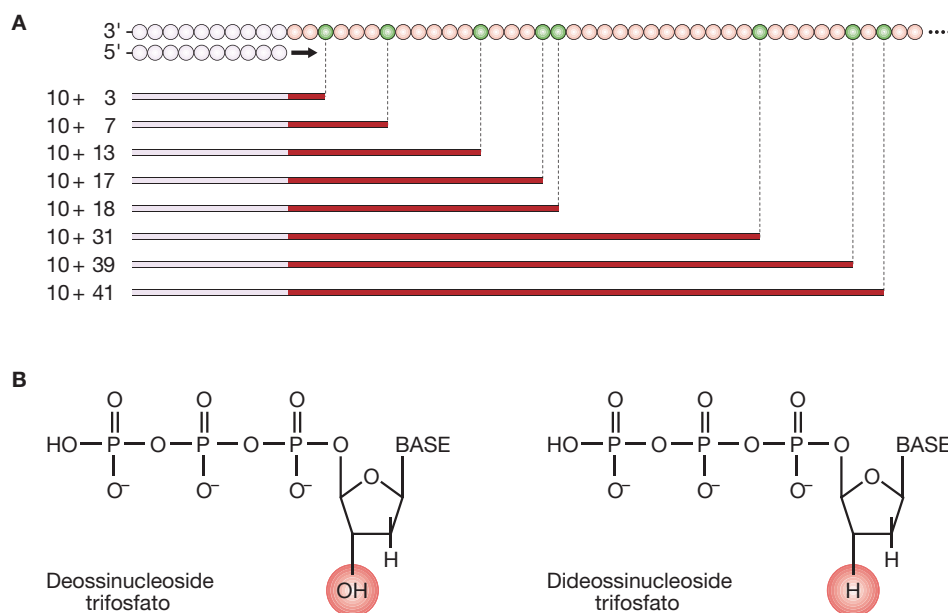


Figura 20.40 Produzione dei frammenti tronchi per il sequenziamento del DNA mediante il metodo enzimatico di Sanger. (A) I frammenti tronchi vengono ottenuti mediante sintesi enzimatica da parte della DNA polimerasi che utilizza come stampo il frammento da sequenziare, clonato in un vettore a singolo filamento, e come innesco un oligonucleotide che viene introdotto nella miscela di reazione e che è complementare a una sequenza nota che

fiancheggia il frammento da sequenziare al 3'. L'interruzione della sintesi in corrispondenza di nucleotidi specifici (in questo caso G, verdi) è dovuta alla presenza nella miscela, insieme ai 4 normali deossiribonucleotidi, di una piccola quantità di un dideossiribonucleotide (in questo caso dideoossi-G). (B) Confronto della struttura di un deossiribonucleotide e di un dideossiribonucleotide: quest'ultimo manca del 3'-OH necessario per l'allungamento della sintesi.

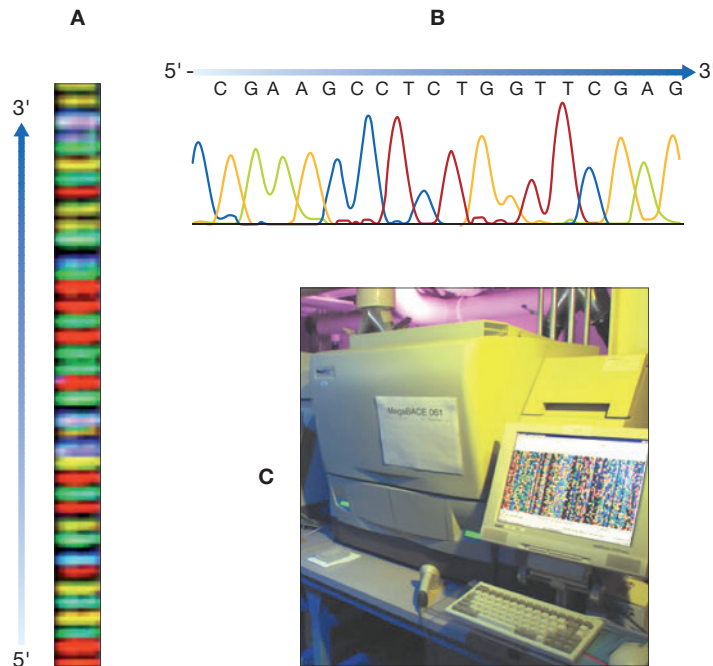


Figura 20.41 Automazione del sequenziamento del DNA. (A) I frammenti tronchi prodotti mediante il metodo enzimatico di Sanger con dideossinucleotidi marcati con quattro fluorofori differenti producono bande di diverso colore per i quattro nucleotidi in un'unica corsia del gel. (B) Le bande di diverso colore corrispondenti ai quattro nucleotidi vengono lette da un rivelatore che invia automaticamente i dati a un computer. (C) Un sequenziatore automatico di DNA.

è radioattivo) e, dopo elettroforesi, messi in evidenza mediante autoradiografia. Il metodo enzimatico di Sanger veniva inizialmente utilizzato meno del metodo chimico, perché meno accurato e inoltre perché presenta la limitazione che si deve in partenza già conoscere la sequenza nucleotidica della regione a cui fare appaiare l'innesco. Col passare degli anni, però, il metodo di Sanger è stato affinato, reso più accurato e, soprattutto, si è rivelato molto più adatto per un'automatizzazione della procedura rispetto al metodo chimico. Infatti, a partire dagli anni novanta un succedersi di innovazioni tecnologiche ha permesso di rendere il sequenziamento quasi completamente automatico ed estremamente rapido abbassando i costi della procedura di vari ordini di grandezza. Una prima innovazione è stata quella di utilizzare nelle reazioni di sintesi non precursori radioattivi ma fluorescenti e, quindi, di marcare con diversi colori i frammenti tronchi che terminano con i quattro nucleotidi. Ciò permette di analizzare mediante elettroforesi i prodotti delle quattro reazioni mescolati, in un'unica corsia del gel. Ancora meglio, se a essere marcati con i diversi fluorofori sono i dideossinucleotidi che terminano le reazioni specificamente per ciascuna delle quattro basi, anche la reazione di sintesi può essere portata avanti in un unico campione, semplificando ulteriormente la procedura. In ogni caso, la sequenza nucleotidica, anziché essere rappresentata su un gel da quattro corsie corrispondenti ai quattro nucleotidi come avevamo visto in **figura 20.38C**, è ora rappresentata da un'unica successione continua di bande di colore diverso per ciascun nucleotide, come mostrato in **figura 20.41A**. Un gel di questo genere viene facilmente letto automaticamente da uno strumento che rileva le bande al loro passaggio attraverso un lettore ottico e invia i dati della sequenza direttamente a un computer (**fig. 20.41B e C**). Un'altra innovazione è l'uso di capillari molto sottili riempiti con il gel di poliacrilammide in cui effettuare l'elettroforesi. Un singolo strumento può gestire, con una risoluzione molto elevata, molte decine di capillari.

20.8 Piattaforme di sequenziamento di nuova generazione

L'evoluzione delle ricerche in Biologia Molecolare richiede sempre di più la conoscenza di sequenze di DNA e delle loro eventuali variazioni correlate alla variabilità tra gli individui di una specie o a specifiche patologie.

Negli ultimi anni le metodologie e le piattaforme tecnologiche per il sequenziamento degli acidi nucleici hanno subito una straordinaria evoluzione che sta rivoluzionando le possibilità e gli orizzonti applicativi in moltissimi settori. L'introduzione sul mercato, a partire dal 2005, delle cosiddette **piattaforme di sequenziamento di nuova generazione** (NGS, *Next Generation Sequencing*) ha prodotto un eccezionale aumento della capacità di sequenziamento, in parallelo a una drastica riduzione dei costi. Basti pensare che con le attuali tecnologie è possibile produrre in un solo giorno lo stesso volume di dati prodotto complessivamente negli ultimi trent'anni attraverso il sequenziamento basato sul metodo di Sanger. Il traguardo di ottenere l'intero genoma umano attraverso una semplice analisi di routine e al costo di soli 1000 \$ è stato raggiunto, come la possibilità di immagazzinarlo, insieme ad altri dati personali, su un chip della tessera sanitaria individuale in modo da consentire la somministrazione di cure personalizzate sulla base del proprio profilo genetico. Questa rivoluzione tecnologica è stata resa possibile grazie al recente sviluppo delle nanotecnologie e di tecniche chimiche più avanzate.

Tra le principali innovazioni introdotte nelle piattaforme NGS le più rilevanti sono che: (1) non è richiesto il clonaggio dei frammenti di DNA da sequenziare, (2) l'utilizzo di micro- o nano-reattori, normalmente immobilizzati su un supporto solido, permette un elevatissimo livello di parallelizzazione e (3) la determinazione della sequenza non richiede il passaggio limitante della separazione elettroforetica in quanto via via che i nucleotidi vengono incorporati nella reazione di sequenziamento vengono simultaneamente identificati.

Normalmente si parte da una preparazione di acidi nucleici di buona qualità. Il DNA viene quindi frammentato con approcci chimici o enzimatici in frammenti di dimensioni appropriate, in funzione della tecnologia usata e dell'applicazione di interesse, ottenendo librerie di frammenti alle cui estremità vengono legati covalentemente degli adattatori che sono poi usati sia per l'amplificazione clonale dei frammenti che per la reazione di sequenziamento. L'amplificazione clonale dei frammenti si rende necessaria per ottenere un segnale misurabile (ad es. intensità di luce, differenza di pH ecc.) che consente poi di determinare la sequenza. Più recentemente, per l'avvento di sistemi di rilevazione ultrasensibili, stanno emergendo nuove tecnologie, dette di "terza generazione", che sono in grado di sequenziare singole molecole di DNA senza la necessità della fase preliminare dell'amplificazione clonale. Questa, infatti, può introdurre delle distorsioni a causa della possibilità di errori nella fase di amplificazione e della non uniformità nella scelta degli stampi (si osserva che alcuni frammenti vengono preferenzialmente amplificati rispetto ad altri in funzione delle loro caratteristiche intrinseche quale, per esempio, il contenuto di G+C). Infine, l'amplificazione clonale preclude la possibilità di rilevare specifiche modificazioni del DNA, come i diversi tipi di metilazione (vedi par. 17.1). Tuttavia, i sistemi di terza generazione, come la piattaforma *Pacific Bioscience* o il piccolo dispositivo realizzato da *Oxford Nanopore*, si stanno diffondendo sempre più e potrebbero soppiantare in breve le piattaforme di seconda generazione se continuerà il loro progressivo sviluppo con il miglioramento del livello di accuratezza e la riduzione dei costi di esercizio.

Attualmente, le piattaforme di sequenziamento di seconda generazione più diffuse prevedono una fase di amplificazione clonale dei frammenti. Esse sono: (1) *454 Genome Sequencer FLX* di Roche; (2) *Genome Analyzer* di Illumina; (3) *SOLiD* di ThermoFisher Scientific; (4) *Ion Torrent/Proton* di ThermoFisher Scientific.

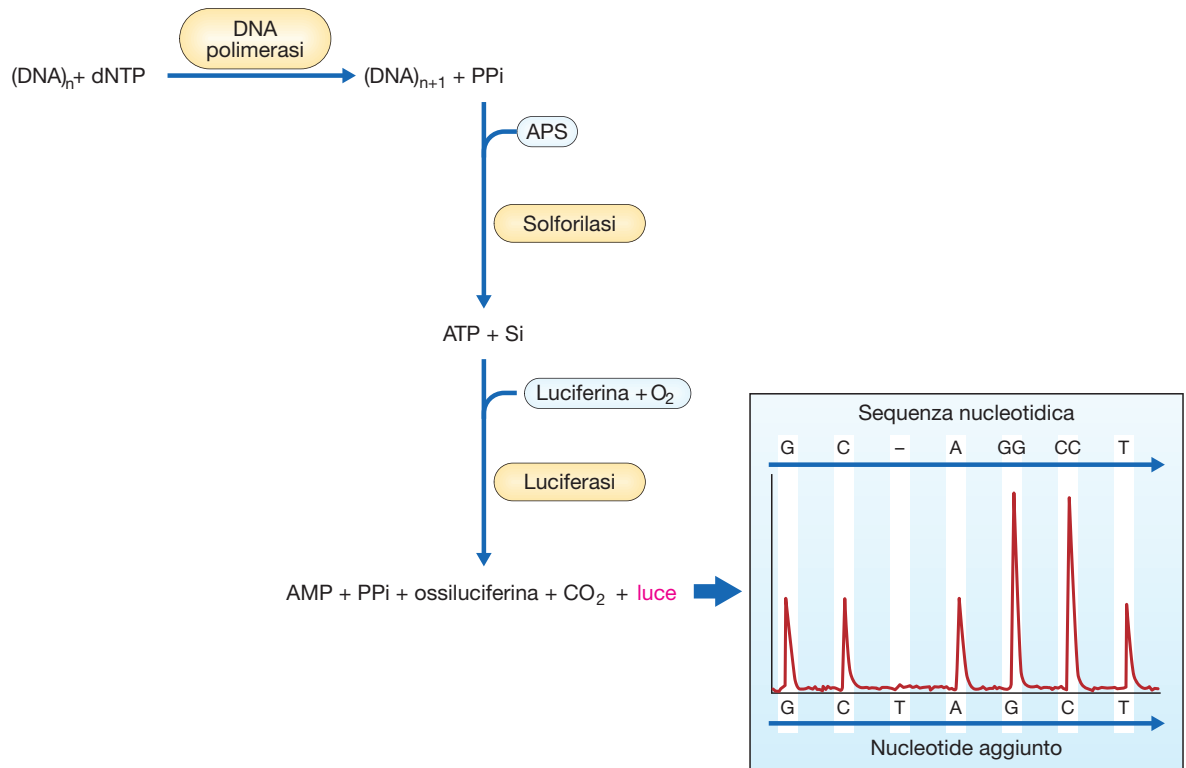
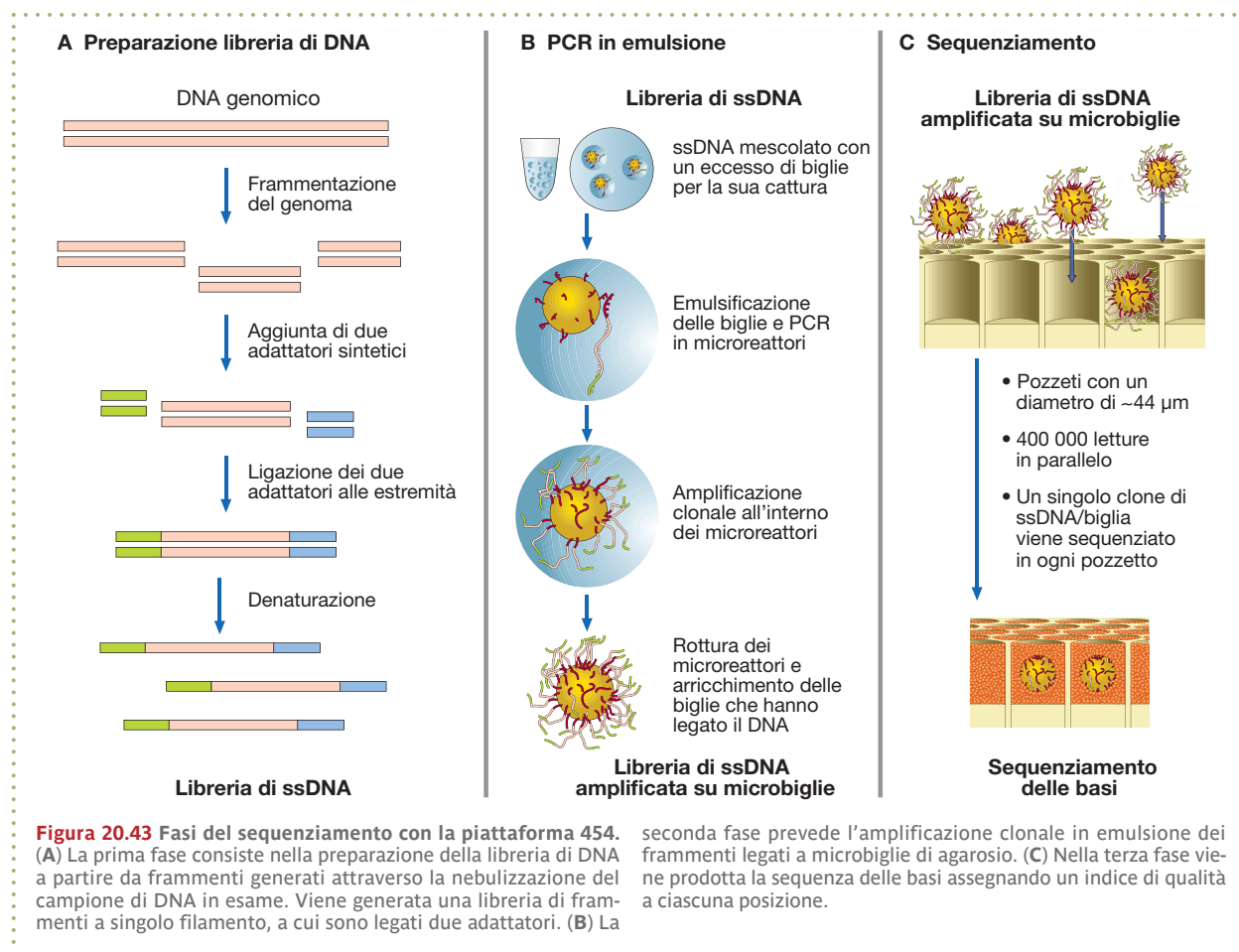


Figura 20.42 Chimica della reazione del pirosequenziatore 454. La DNA polimerasi catalizza l'incorporazione del deossinucleoside trifosfato al filamento di DNA se è complementare alla base del filamento stampo. Ogni evento di incorporazione è accompagnato dal rilascio di pirofosfato (PPi) in quantità equimolare a quella del nucleotide incorporato. In presenza di adenosina 5' fosfosolfato (APS), l'ATP solforilasi converte quantitativamente il PPi ad ATP, che, a sua volta, guida la conversione, catalizzata dalla luciferasi, di luciferina a ossiluciferina con conseguente produzione di luce di

intensità proporzionale alla quantità di ATP. La luce prodotta è rilevata da una CCD camera e visualizzata come picco in un pirogramma. A ogni ciclo della reazione vengono aggiunti sequenzialmente i quattro deossinucleotidi. I deossinucleotidi in eccesso che non vengono incorporati, vengono degradati dall'enzima apirasi. Poiché il dATP è un substrato naturale della luciferasi, nella reazione di polimerizzazione viene usato un suo analogo, la deossiadenosina α -tiorifosfato (dATPS), che viene utilizzata efficientemente dalla DNA polimerasi ma non viene riconosciuta dalla luciferasi.

Di seguito si riporta una sommaria descrizione delle diverse piattaforme sopra citate, evidenziandone limiti e potenzialità.

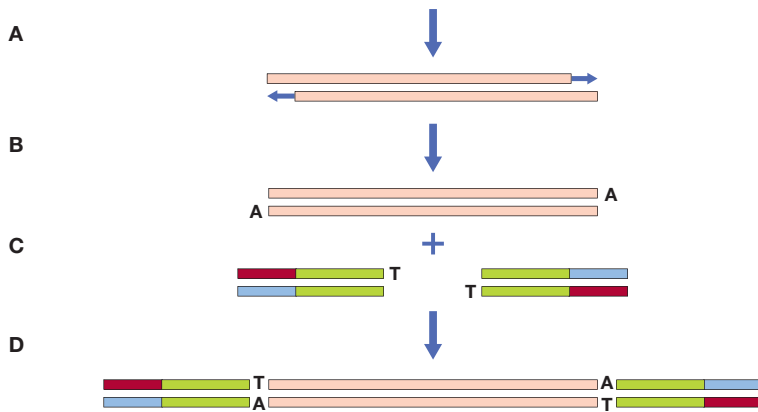
La **tecnologia 454** è basata sulla rilevazione della chemiluminescenza prodotta dal rilascio del pirofosfato nella reazione di sintesi del DNA (**fig. 20.42**). Il processo di sequenziamento avviene in tre fasi (**fig. 20.43**). In particolare, è critica la fase della PCR in emulsione, nella quale si utilizzano microbiglie di agarosio a cui sono legate singole molecole di DNA che vengono così amplificate in modo clonale nell'emulsione acqua-olio costituita da micelle contenenti tutti i reagenti necessari per la reazione di PCR. L'amplificazione clonale genera circa un milione di molecole identiche sulla superficie di ciascuna microbiglia di agarosio. Le microbiglie così ottenute vengono depositate su un apposito vetrino dove vengono alloggiare singolarmente in micropozzetti di dimensioni appropriate. Vengono quindi aggiunti gli enzimi e i reagenti per la reazione di pirosequenziamento attraverso un apposito apparato di microfluidica. Nel corso della reazione di sequenziamento una camera CCD registra la luce emessa da ciascun micropozzetto. L'intensità luminosa sarà proporzionale al numero di basi incorporate in ciascun ciclo. Per esempio, l'incorporazione di 2 A produrrà un'emissione di intensità circa doppia rispetto all'incorporazione di una singola A. La linearità tra intensità luminosa e numero di basi incorporate degrada in funzione della lunghezza dell'omopolimero e sono possibili errori oltre 5-6. Un singolo esperimento produce circa un milione di letture di sequenza (*read*) che possono raggiungere la lunghezza di 1000 pb.



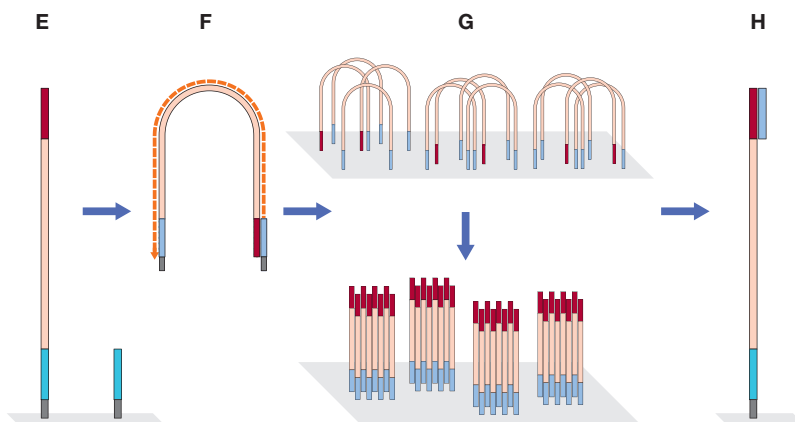
seconda fase prevede l'amplificazione clonale in emulsione dei frammenti legati a microbiglie di agarosio. (C) Nella terza fase viene prodotta la sequenza delle basi assegnando un indice di qualità a ciascuna posizione.

Anche la **piattaforma Illumina** prevede una fase di amplificazione clonale del DNA ottenuto dalla frammentazione del campione in esame. I frammenti di DNA ottenuti, di lunghezza pari a 500-700 pb, opportunamente modificati dall'aggiunta di una coppia di adattatori alle due estremità, vengono fissati a un vetrino sul quale erano stati precedentemente legati un gran numero di oligonucleotidi complementari agli adattatori (**fig. 20.44**). È così possibile realizzare un processo di amplificazione a ponte (*bridge amplification*) che genera cluster clonali dei frammenti, ciascuno costituito da circa un milione di copie. A questo punto può aver luogo il sequenziamento associato alla sintesi del DNA. In questa fase, dopo l'appaiamento di un primer appropriato i quattro deossinucleotidi vengono aggiunti simultaneamente. Ciascun deossinucleotide è marcato con un fluorocromo diverso e porta un gruppo bloccante al 3'-OH che impedisce l'incorporazione di ulteriori nucleotidi. Una volta rilevata la fluorescenza viene rimosso il blocco dal 3'-OH ed eseguito un nuovo ciclo di sequenziamento. L'uso di inibitori reversibili rende possibile la lettura di una base per volta e non presenta i problemi legati alla determinazione della sequenza di tratti omopolimerici come nel caso del pirosequenziamento. Tra le peculiarità delle piattaforme Illumina vi è la possibilità di ottenere le sequenze di entrambe le estremità dei frammenti ottenuti (*paired-end sequencing*), rendendo quindi molto più accurata e specifica l'analisi bioinformatica dei dati. Lo strumento Illumina più avanzato, che ha raggiunto e superato il traguardo di un genoma umano al costo di 1000 \$, è in grado di produrre 6000 Gpb e 20 miliardi di *read* in un singolo esperimento, con letture paired-end che possono raggiungere la lunghezza di 300 pb per ciascuna estremità del frammento.

Preparazione libreria



Generazione di cluster clonali



Sequenziamento

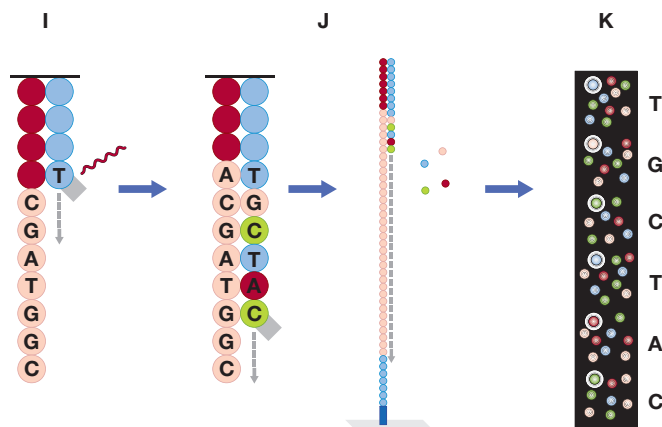


Figura 20.44 Rappresentazione schematica delle fasi del sequenziamento con il sistema Illumina.

La prima fase consiste nella preparazione della libreria di frammenti. Le estremità dei frammenti vengono completate (A); ad entrambe le estremità si aggiunge una adenina che protrude dalla doppia elica (B). Vengono quindi legati gli adattatori (C e D) e si effettua la selezione dei frammenti di DNA cui questi sono legati. Nella seconda fase il DNA viene fissato sulla superficie di una cella a flusso (E) e si effettua una amplificazione a ponte (F, *bridge amplification*). L'amplificazione genera cluster clonali di frammenti (G) cui vengono poi appaiati i primer per il sequenziamento (H). Nella terza fase il sequenziamento avviene aggiungendo un deossinucleotide per volta e rilevandone l'emissione fluorescente (I). Dai deossinucleotidi incorporati viene quindi rimosso il gruppo bloccante, e il processo continua con l'aggiunta di altri deossinucleotidi (J). Il ciclo viene così ripetuto un numero prefissato di volte, leggendo ogni volta la fluorescenza (K). In questo modo tutte le sequenze ottenute avranno la stessa lunghezza.

Il **sistema SOLiD** utilizza una libreria di frammenti modificati con adattatori come le altre due piattaforme precedentemente descritte e richiede una fase di amplificazione clonale che può essere condotta secondo due modalità: in emulsione analogamente a quanto già descritto per il sistema 454 o, nelle versioni più recenti, su vetrino utilizzando una metodologia simile a quella adottata da Illumina, denominata "Wildfire". La peculiarità di questo sistema sta nell'utilizzo di una reazione di ligazione per effettuare il sequenziamento (*sequencing by ligation*, **fig. 20.45**).

Le microbiglie con gli stampi di DNA a singolo filamento, amplificati in modo clonale, vengono depositate su un vetrino di una cella a flusso, ovvero, qualora l'amplificazione clonale sia condotta con la metodologia Wildfire, viene direttamente utilizzata la cella a flusso su cui è stata condotta l'amplificazione clonale. Il sequenziamento viene avviato dall'appaiamento di un primer complementare all'adattatore. Il primer è orientato in modo da offrire un'estremità 5' fosfato per la reazione di ligazione. Viene quindi utilizzata una serie di sonde dette "di interrogazione" costituite da una libreria di ottameri, marcati con quattro diversi fluorocromi in funzione dell'identità dei primi due nucleotidi. Disponendo di 4 colori per marcare 16 diversi dinucleotidi, avremo che uno stesso colore corrisponderà a 4 dinucleotidi differenti, secondo lo schema illustrato in **figura 20.45**. La libreria di ottameri è costruita in modo tale che le prime 5 basi si appaiano in modo specifico mentre le ultime 3, costituite da inosina, in modo aspecifico.

Dopo l'appaiamento della sonda specifica e la reazione di ligazione del primo ottamero, tutti gli ottameri non appaiati vengono rimossi con un lavaggio. Viene rilevata la fluorescenza e successivamente, utilizzando ioni argento, vengono staccati gli ultimi tre nucleotidi della sonda ottamerica legati attraverso un legame fosforotioato. Lo stesso ciclo sopra descritto viene ripetuto più volte, estendendo il primer originariamente appaiato. In questo modo viene rilevato il "colore" relativo ai dinucleotidi 1-2, 6-7, 11-12 ecc. Il numero di cicli che vengono eseguiti è di norma pari a sette, ma le versioni più aggiornate del sistema permettono un numero di cicli maggiore. Dopo l'ultimo ciclo, il filamento ottenuto dalle ligazioni successive viene rimosso per denaturazione e si riparte per un nuovo round di cicli con un nuovo primer sfasato di un nucleotide rispetto al precedente ($n - 1$). In questo modo verrà rilevato il colore relativo alle posizioni 0-1, 5-6, 10-11 ecc. Vengono poi eseguiti altri tre round con i primer $n - 2$, $n - 3$ e $n - 4$ in modo da leggere in doppio il colore associato a ciascuna posizione. Si rammenti che ciascuno dei quattro colori corrisponde a quattro differenti dinucleotidi, secondo lo schema in **figura 20.45**. Per la traduzione o deconvoluzione della "sequenza a colori" nella sequenza nucleotidica è sufficiente conoscere almeno una delle basi, come quella in posizione 0, che corrisponde all'ultima posizione dell'adattatore (**fig. 20.46**). È evidente che nota la prima base della sequenza di colori, essa può essere univocamente decifrata. È immediato verificare che una mutazione nella sequenza comporta il

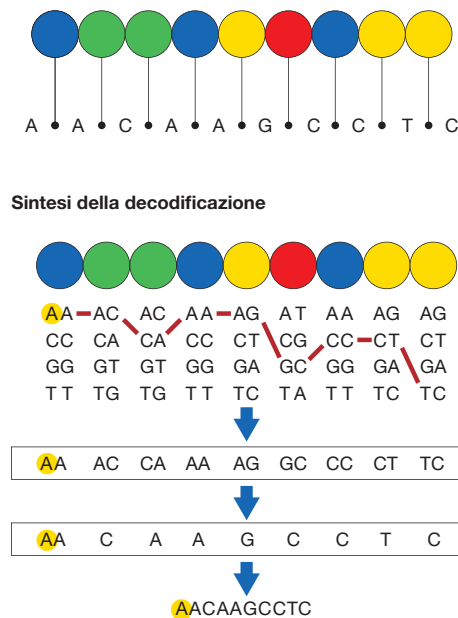
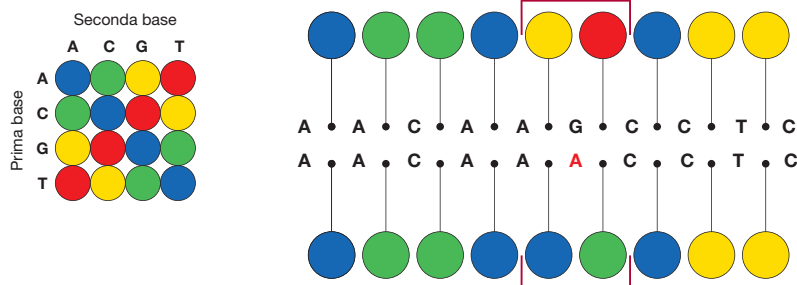


Figura 20.46 Interpretazione della "sequenza a colori" prodotta dal sequenziatore SOLiD. Nella sequenza a colori ciascun dinucleotide corrisponde a un colore. Nell'esempio mostrato avremo quindi una sequenza di 9 colori corrispondenti a 10 nucleotidi. Per la deconvoluzione della sequenza di colori in sequenza di nucleotidi è sufficiente conoscere l'identità del primo nucleotide nel primo "colore" (nell'esempio la A evidenziata in giallo corrispondente all'ultima posizione dell'adattatore).

Figura 20.47 Una mutazione nella sequenza comporta il cambiamento di due colori adiacenti.



cambiamento di due colori adiacenti (fig. 20.47). Questa proprietà è certamente vantaggiosa per distinguere gli errori di sequenziamento dalle reali mutazioni della sequenza. Le piattaforme di sequenziamento SOLiD più recenti sono in grado di generare circa 320 Gpb di letture di lunghezza fino a 75 pb.

La tecnologia utilizzata dal sistema Ion Torrent/Proton/S5 di ThermoFisher Scientific è molto simile a quella del sistema 454. In questo caso si sfrutta il principio per cui per ciascun nucleotide incorporato viene rilasciato un protone (H^+) dal gruppo 3'-OH coinvolto nella formazione del legame fosfodiesterico. La reazione di polimerizzazione avviene su un chip in grado di alloggiare milioni di biglie sulla cui superficie i frammenti di DNA hanno subito un processo di amplificazione clonale in modo analogo a quanto descritto per la piattaforma 454. Durante la reazione di sequenziamento il sistema elettronico collegato al chip è in grado di rilevare variazioni di pH che sono proporzionali al numero di nucleotidi incorporati sequenzialmente. La peculiarità di questa tecnologia rispetto alle altre descritte in precedenza sta nella maggiore semplicità, visto che non è richiesto un apparato ottico, normalmente più complesso e costoso, per la rilevazione del segnale. Questa piattaforma è in grado di generare fino a 60-80 milioni di letture di lunghezza fino a 600 nt.

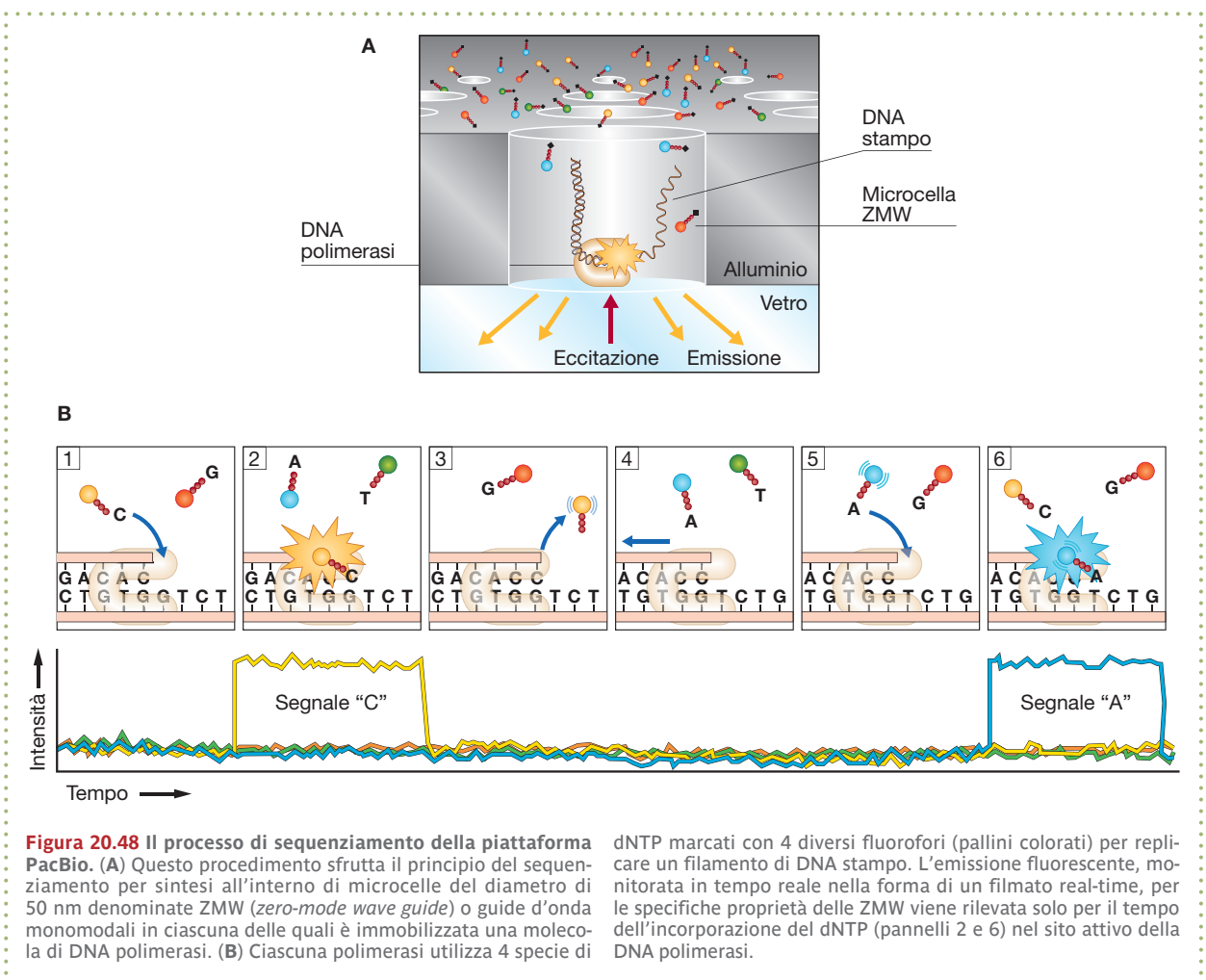
Tra le tecnologie precedentemente descritte, il sistema Illumina è stato oggetto di un maggiore e progressivo sviluppo che gli ha consentito di guadagnare la leadership nel mercato, mentre i sistemi Roche e SOLiD sono stati via via abbandonati. Tuttavia, i cosiddetti **sistemi di sequenziamento di terza generazione** stanno facendo grandissimi progressi e non è escluso che possano del tutto soppiantare i sistemi NGS già descritti.

Infatti, i sistemi di sequenziamento di seconda generazione descritti in precedenza hanno due punti deboli: 1) la necessità dell'amplificazione clonale dei frammenti da sequenziare per rendere possibile l'emissione di un segnale rilevabile per l'identificazione dei nucleotidi incorporati; 2) le ridotte dimensioni delle *read* prodotte.

L'amplificazione clonale, oltre a essere una fase piuttosto laboriosa in diversi sistemi di sequenziamento NGS, può introdurre degli errori sistematici che si riflettono nella eterogeneità del "coverage" del sequenziamento, pari al numero di *read* che si allineano in una specifica posizione, con alcune regioni con coverage molto alto e altre che non vengono sequenziate affatto. Inoltre, più le *read* sono corte più difficoltoso è il loro "assemblaggio" per ricostruire la sequenza consenso. Per queste ragioni, c'è crescente interesse nei sistemi di sequenziamento detti di "terza generazione" che da un lato sono in grado di effettuare il sequenziamento su singola molecola di DNA senza la necessità di una preliminare amplificazione clonale e dall'altro sono in grado di generare *read* di lunghezza significativamente maggiore rispetto agli apparati di seconda generazione. I sistemi Pacific Bioscience e Oxford Nanopore sono le due piattaforme di sequenziamento di terza generazione al momento disponibili per i ricercatori.

■ Pacific Bioscience

La piattaforma di sequenziamento PacBio RS, rilasciata dalla Pacific Biosciences of California, Inc (PacBio), costituisce il primo esempio di tecnica di sequenziamento del DNA a singola molecola. La tecnologia di sequenziamento sviluppata da PacBio, è stata commercializzata a partire dal 2010, ed è più comunemente nota con l'acronimo SMRT, *single-molecule real-time sequencing*. Questa tecnologia si avvale di sofisticati strumenti ottici, le guide d'onda monomodali (ZMW, *zero-mode waveguide*), che sono strutture in grado di permettere una propagazione guidata delle onde elettromagnetiche così da incrementare in modo significativo la sensibilità di lettura, consentendo di rilevare il segnale di fluorescenza emesso dall'incorporazione di un singolo dNTP nel corso dell'estensione del filamento di DNA. Per questa ragione il sequenziamento può avvenire a livello di singola molecola senza la necessità di amplificare il DNA di partenza tramite PCR. Lo speciale chip di sequenziamento sviluppato da PacBio consiste in un array di guide d'onda da 50 nm (SMRT cell) alle cui estremità sono legate delle DNA polimerasi (**fig. 20.48**). Ogni SMRT cell, che contiene 150 000 ZMW, è in grado di generare circa 80 000 *read* di buona qualità con lunghezza media di 10-15 kpb corrispondenti a 1 Gpb totali. Dato che il filmato capta l'informazione relativa alla velocità di incorporazione dei nucleotidi, è anche possibile determinare le eventuali modificazioni presenti sugli stessi (ad es. N^6 -metil-adenina, N^4 -metil-citosina, 5-metil-citosina, ecc.) che alterano in modo sistematico la velocità di incorporazione.



dNTP marcati con 4 diversi fluorofori (pallini colorati) per replicare un filamento di DNA stampo. L'emissione fluorescente, monitorata in tempo reale nella forma di un filmato real-time, per le specifiche proprietà delle ZMW viene rilevata solo per il tempo dell'incorporazione del dNTP (pannelli 2 e 6) nel sito attivo della DNA polimerasi.

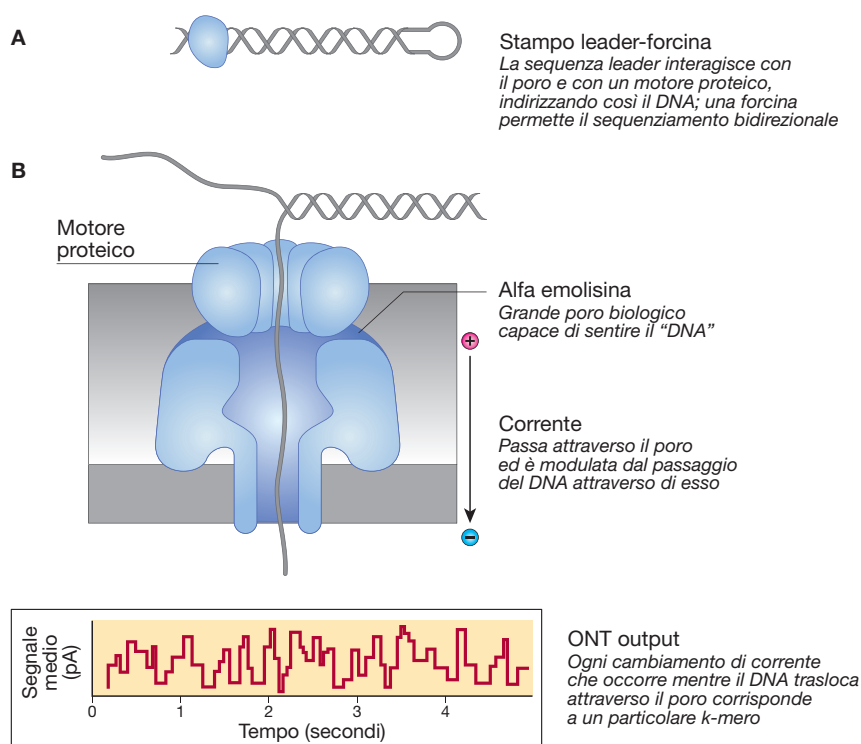
Il tasso di errore si attesta attorno al 19%. Gli errori prevalenti sono costituiti da inserzioni e delezioni di base, causati probabilmente da incorporazioni multiple troppo veloci, che non vengono riconosciute in maniera appropriata dal sistema. Nonostante l'alto tasso di errore uno dei vantaggi maggiori di questa tecnologia consiste nel fatto che gli errori di sequenziamento sembrano essere distribuiti in maniera casuale e non occorrono preferenzialmente in sequenze con una predeterminata composizione. Se il sequenziamento viene eseguito a una profondità adeguata, è possibile utilizzare opportuni algoritmi (tra cui alcuni proposti sviluppati dalla PacBio stessa) per allineare tra loro le sequenze ottenute e ridurre drasticamente il tasso di errore, arrivando a un'accuratezza che può sfiorare il 99,9%, con un coverage $>30\times$.

■ Oxford Nanopore

Il sistema Oxford Nanopore Technology (ONT) utilizza un approccio molto innovativo per cui la lettura della sequenza nucleotidica viene effettuata in funzione dell'alterazione di corrente misurata in seguito all'attraversamento di un singolo filamento di DNA attraverso un nanoporo proteico (ad es. alfa-emolisina). Al DNA da sequenziare viene legata un'altra proteina che regola la velocità di attraversamento del poro e consente la lettura di uno (lettura 1D) o entrambi (lettura 2D) i filamenti complementari di DNA (**fig. 20.49**). Il passaggio del DNA attraverso il nanoporo causa una perturbazione nell'intensità di corrente che è caratteristica e distinguibile per ciascun nucleotide. Misurando queste perturbazioni è possibile determinare la sequenza di una molecola di DNA. Dato che il sequenziamento prevede interazioni di tipo 1:1 tra ciascun nanoporo e una molecola di DNA, questa tecnologia è in grado di sequenziare DNA a singola molecola e non richiede che il DNA da sequenziare venga amplificato per PCR. Un'altra caratteristica importante di questo sistema è che, come PacBio, anche ONT consente di determinare direttamente modificazioni nucleotidiche durante il sequenziamento.

Figura 20.49 La piattaforma Oxford Nanopore Technology (ONT).

(A) Al DNA, frammentato a 8-10 kpb, vengono legati due differenti adattatori, un leader e un hairpin. (B) L'adattatore leader indirizza il frammento verso il poro interagendo con una proteina motore (in giallo), mentre l'adattatore hairpin consente il sequenziamento bidirezionale. La traslocazione del filamento di DNA attraverso il poro genera un'alterazione del voltaggio specifica della sequenza delle basi. (Adattato da Goodwin, *Nat Rev Genet.* 2016.)



Inoltre è stato annunciato un protocollo per il sequenziamento diretto dell'RNA che faciliterà in modo significativo lo studio dell'RNA editing.

Una caratteristica molto vantaggiosa di questo sistema consiste nel fatto che l'efficienza dei nanopori degrada più lentamente di quella di una polimerasi ed è perciò possibile leggere sequenze molto più lunghe in una singola reazione. Le problematiche principali legate a questa tecnologia consistono nel fatto che, con le attuali tecnologie, risulta praticamente impossibile associare una singola misurazione di variazione della corrente elettrica al passaggio di una singola base in un nanoporo. Un altro importante limite deriva dal fatto che la sequenza viene letta in maniera indiretta e completamente slegata da qualsiasi reazione biochimica. L'algoritmo per determinare le sequenze, utilizzato con questa tecnologia, adotta tecniche di apprendimento automatico basate sul "machine learning": tramite opportune procedure, il software viene addestrato a riconoscere i diversi segnali sulla base di analisi di sequenze note e in seguito a discriminare e riconoscere sequenze diverse. Questa procedura può comportare problemi di accuratezza nella chiamata delle basi, specialmente nel caso in cui in un futuro la macchina dovesse analizzare una sequenza con composizione in GC molto diversa rispetto a quella del/dei genomi sui quali è avvenuto l'addestramento.

Per quanto questa tecnologia sia ancora relativamente giovane, gli studi pilota condotti sui primi dati di sequenziamento di questo tipo evidenziano che il tasso di errore del sequenziamento risulta piuttosto alto con una media del 20%, e mostra una proporzione circa eguale di inserzioni, delezioni e sostituzioni di base. Se si raggiunge una profondità di sequenziamento adeguata, è però possibile utilizzare opportuni algoritmi di correzione degli errori, basati sull'allineamento delle *read* ottenute per ridurre il tasso di errore. Si stima che in questo modo il tasso di errore possa essere ridotto attorno al 3-5%.

Lo strumento attualmente più diffuso, denominato Minion, è di dimensioni molto contenute con un peso di soli 90 g e viene collegato a un PC con un semplice cavo USB. Con i sistemi ONT di ultima generazione è oggi possibile ottenere *read* della lunghezza media di 4-5 kpb, per complessivi 5-10 Gpb per run, e accuratezza >96%.

L'avvento delle piattaforme NGS ha prodotto un'enorme accelerazione in moltissimi ambiti della ricerca genomica. I principali campi di applicazione delle piattaforme NGS sono:

- sequenziamento *de novo* e ri-sequenziamento di genomi completi;
- identificazione di siti polimorfici e mutazioni;
- analisi su larga scala del trascrittoma;
- mappaggio su scala genomica di siti di interazione tra DNA (RNA) e proteine e analisi epigenetiche dell'intero genoma;
- caratterizzazione del metiloma e studio dell'editing dell'RNA;
- analisi metagenomiche.

■ Sequenziamento *de novo* e ri-sequenziamento di genomi completi

L'elevatissima produttività delle piattaforme NGS rende possibile il ri-sequenziamento di genomi completi, anche molto complessi come quello umano. Infatti, al fine di ottenere una mappa completa della variabilità genetica nell'uomo, nel 2012 è stato completato da un consorzio internazionale un progetto volto al sequenziamento completo di 1000 genomi di individui di differenti provenienze geografiche. La lunghezza limitata delle letture di sequenza che normalmente è possibile ottenere dalle piattaforme NGS non è un problema nei progetti di ri-sequenziamento quando si dispone di un genoma di riferimento, ma è una seria limitazione nel caso di sequenziamento *de novo* di genomi eucariotici complessi. La possibilità di ottenere sequenze da entrambe le estremità di frammenti di DNA di diversa lunghezza (*paired-end sequencing*) o di ottenere *read* molto lunghe

(>5 Kpb), attualmente possibile con alcune delle piattaforme NGS precedentemente descritte, può permettere il sequenziamento genomico *de novo* facilitandone l'assemblaggio. Il sequenziamento paired-end, dato che è nota la dimensione media dei frammenti selezionati, consente anche di determinare variazioni strutturali nel genoma (ad es. estese inserzioni e delezioni) permettendo, per esempio, un'analisi dettagliata delle alterazioni a livello genomico delle cellule tumorali. Oltre alle sequenze paired-end, sono anche disponibili protocolli sperimentali utilizzabili su diverse piattaforme NGS che consentono di ottenere sequenze "mate-pair", ovvero relative alle estremità di frammenti molto più grandi, fino a 10-20 kpb. In questo modo il processo di assemblaggio risulta significativamente facilitato. Infine, una strategia molto efficace per ottimizzare il sequenziamento *de novo* mediante NGS prevede l'utilizzo in parallelo di sistemi di sequenziamento a elevata produttività quale Illumina, insieme a sistemi di sequenziamento in grado di produrre letture molto lunghe come *Pacific Bioscience* od *Oxford Nanopore*.

■ Identificazione di siti polimorfici e mutazioni

L'elevata "profondità" del sequenziamento delle piattaforme NGS consente uno studio su scala genomica della variabilità genetica permettendo l'identificazione di SNP (*Single Nucleotide Polymorphisms*) e mutazioni, anche molto poco frequenti. Queste analisi possono essere condotte sull'intero genoma, oppure focalizzate sulle sue regioni codificanti. In quest'ultimo caso, prima del sequenziamento la libreria di frammenti di DNA subisce un processo di arricchimento mediante ibridazione a sonde disegnate sulle regioni esoniche note del genoma. Questa tecnica, nota come **sequenziamento dell'esoma** (**WES**, *Whole Exome Sequencing*) è stata diffusamente applicata per la caratterizzazione di mutazioni causative di moltissime patologie. Infine, una volta definito il pannello di geni coinvolti in una specifica patologia è anche possibile effettuarne il loro sequenziamento mirato utilizzando specifici sistemi di arricchimento per ibridazione. Numerosi pannelli genici sono attualmente in commercio per l'analisi mutazionale specifica per diverse patologie (ad es. cardiomiopatie).

■ Analisi del trascrittoma

L'analisi su larga scala del trascrittoma (**RNA-Seq**) è fondamentale per determinare a livello qualitativo e quantitativo il piano di espressione genica in diversi tipi cellulari e condizioni e la caratterizzazione delle diverse varianti di splicing. A questo scopo una maggiore lunghezza delle sequenze, come quella che si ottiene con le piattaforme di sequenziamento di terza generazione, è fondamentale per determinare le connessioni tra gli esoni dei diversi trascritti alternativi. Allo stesso modo si possono condurre analisi su larga scala per identificare gli RNA non codificanti, come i piccoli miRNA, e determinarne il profilo di espressione. La metodica RNA-Seq che consente un'analisi "digitale" del profilo di espressione genica ha rapidamente soppiantato le analisi basate sull'utilizzo dei DNA microarray.

■ Studio dell'interazione DNA-proteine

La metodologia **ChIP-Seq**, che prevede il sequenziamento massivo del DNA raccolto in seguito all'immunoprecipitazione della cromatina (*vedi* par. 20.11), consente di determinare una mappa su scala genomica dei siti di legame tra DNA (o RNA) e una o più proteine in esame per cui sia disponibile un anticorpo specifico. Utilizzando anticorpi specifici per determinate modificazioni istoniche (ad es. i siti di metilazione e acetilazione) è possibile ottenere la mappa completa dello stato della cromatina nei tipi cellulari e condizioni sperimentali in esame. Un esempio di esperimento di ChIP-Seq è mostrato nella finestra 11.2.

■ Caratterizzazione del metiloma e studio dell'editing

La mappa dei siti di metilazione sul genoma può essere agevolmente determinata dal sequenziamento massivo del genoma prima e dopo il trattamento con

bisolfito, che converte le citosine non metilate in uracile. Allo stesso modo, confrontando la sequenza genomica e i relativi trascritti è possibile identificare i siti di RNA editing (*vedi* par. 13.2). Le piattaforme di sequenziamento di terza generazione, come abbiamo visto, possono consentire il rilevamento diretto delle modificazioni nucleotidiche in tempo reale nel corso del sequenziamento anche se i protocolli relativi sono ancora in fase di sviluppo e ottimizzazione.

■ Metagenomica

La metagenomica è uno dei campi applicativi più interessanti e promettenti delle tecnologie NGS. Essa consiste nel sequenziamento massivo degli acidi nucleici contenuti in campioni ambientali di diversa provenienza, per esempio di ambienti estremi come regioni aride o polari o in campioni clinici come feci o mucosa intestinale. In questo modo è possibile ottenere un quadro dettagliato della diversità microbica presente nel campione, sia per quanto riguarda gli aspetti tassonomici che funzionali. Se si considera che la gran parte dei microrganismi presenti in natura è tuttora sconosciuta e non coltivabile in laboratorio, si comprendono le enormi potenzialità di questo approccio. La caratterizzazione del microbioma umano, attualmente oggetto di diversi progetti internazionali, può far luce sulle interazioni tra l'uomo e la sua flora microbica e le sue implicazioni nell'origine delle patologie. Inoltre, la caratterizzazione delle funzioni enzimatiche di microrganismi adattati a vivere in ambienti estremi potrebbe contribuire all'identificazione di nuovi enzimi e molecole d'interesse biotecnologico.

L'enorme mole di dati prodotti dalle tecnologie NGS rende ancora più importante la possibilità di disporre di adeguati strumenti bioinformatici per l'immagazzinamento, la gestione e l'analisi dei dati (*vedi* cap. 21) e di infrastrutture hardware adeguate.

20.9 Metodologie recenti per la modificazione mirata dei genomi ("genome editing")

La possibilità di introdurre mutazioni puntiformi in un tratto relativamente breve di DNA è stata individuata alla fine degli anni settanta ed è qui descritta nella **finestra 20.5**. Descriviamo qui, invece, alcune tecniche sviluppate solo recentemente che permettono di modificare in modo mirato un intero genoma.

Nel paragrafo precedente abbiamo descritto nuove piattaforme di sequenziamento del DNA che hanno permesso di identificare la sequenza di centinaia di genomi di procarioti ed eucarioti, sia animali che vegetali. In parallelo a questa rivoluzione tecnologica sono state sviluppate nuove metodologie che, introducendo modificazioni mirate nei genomi, possono causare l'inattivazione di un gene o la generazione di nuove funzioni genetiche con potenziali ed enormi applicazioni in biologia, medicina e agricoltura.

Queste tecniche innovative si basano tutte sull'utilizzo di **nucleasi sequenza-specifiche** che incidono il DNA in punti precisi generando, di solito, rotture su entrambi i filamenti (chiamate **DSBs**, *double-strand breaks*) che, come abbiamo visto nel capitolo 7, sono riparate mediante due meccanismi ricombinativi, chiamati **NHEJ** (*Non-Homologous End Joining*) e **HR** (*Homologous Recombination*). La riparazione tramite NHEJ è spesso associata alla formazione di corte inserzioni e/o delezioni di nucleotidi (*indels*) che ne causano, solitamente, l'inattivazione (**knock-out**). La riparazione tramite HR, invece, avviene normalmente tramite ricombinazione con le sequenze omologhe presenti sul cromosoma non danneggiato (*vedi* fig. 7.10 e cap. 8). Tuttavia, la ricombinazione può anche avvenire con molecole di DNA introdotte sperimentalmente all'interno delle cellule. In tal modo è tecnicamente possibile generare specifiche mutazioni, ma anche cor-

FINESTRA 20.4

IL NORTHERN BLOT E IL WESTERN BLOT

Come abbiamo descritto nella finestra 20.3, molecole di DNA, RNA e proteine di diverse dimensioni possono essere separate mediante elettroforesi su gel. Dopo elettroforesi, sia gli acidi nucleici che le proteine possono essere (se presenti in quantità sufficienti) visualizzati mediante opportune colorazioni. Come indicato nel testo descrivendo il Southern blotting (*vedi* fig. 20.33), però, tali molecole sono quasi sempre presenti in quantità relative non evidenziabili come bande distinte. Per la loro evidenziazione si ricorre, pertanto, a **tecniche di blotting**, cioè al loro trasferimento su filtri di nitrocellulosa e alla successiva incubazione di tali filtri con opportune sonde che possono ibridare con le molecole d'interesse. Per evidenziare molecole di RNA si ricorre alla tecnica chiamata **Northern blot**: tale terminologia è stata adottata mimando la nomenclatura del Southern blot, con la differenza che, in questo caso, non esiste alcun "Signor Northern"! Molecole di RNA di varie dimensioni (ad es. i diversi mRNA presenti in un tessuto) sono separate mediante elettroforesi in gel di agarosio e trasferite su un foglio di nitrocellulosa, in modo del tutto analogo a quanto descritto per il Southern blot nel testo. Il filtro viene poi incubato con una sonda di DNA radioattivo (**fig. 20F4.1A**) che può ibridare

con una molecola di RNA con sequenza complementare alla sonda: nel Northern blot, cioè, si utilizza un'ibridazione RNA-DNA per evidenziare specifiche molecole di mRNA. La procedura di visualizzazione delle molecole che hanno ibridato con la sonda è identica a quella descritta per il Southern blot (*vedi* fig. 20.33).

Miscele di proteine con diverse dimensioni possono essere separate tramite elettroforesi SDS-PAGE e colorate con Coomassie blue (**fig. 20F4.1B**). Spesso, però, la miscela di proteine è composta da specie proteiche così numerose che le bande colorate con il Coomassie blue sfumano una nell'altra. È, però, possibile evidenziare inequivocabilmente la proteina d'interesse nella miscela se si dispone di un anticorpo contro tale proteina. Infatti, le proteine separate mediante SDS-PAGE sono trasferite elettricamente su un foglio di nitrocellulosa (**Western blot**).

Il foglio è poi incubato in presenza dell'anticorpo specifico contro la proteina d'interesse e il complesso proteina-anticorpo formatosi sul gel è evidenziato utilizzando un anticorpo secondario marcato, o in grado di conferire una colorazione specifica, o una chemiluminescenza in grado di dare un segnale su una lastra fotosensibile (**fig. 20F4.1C**).

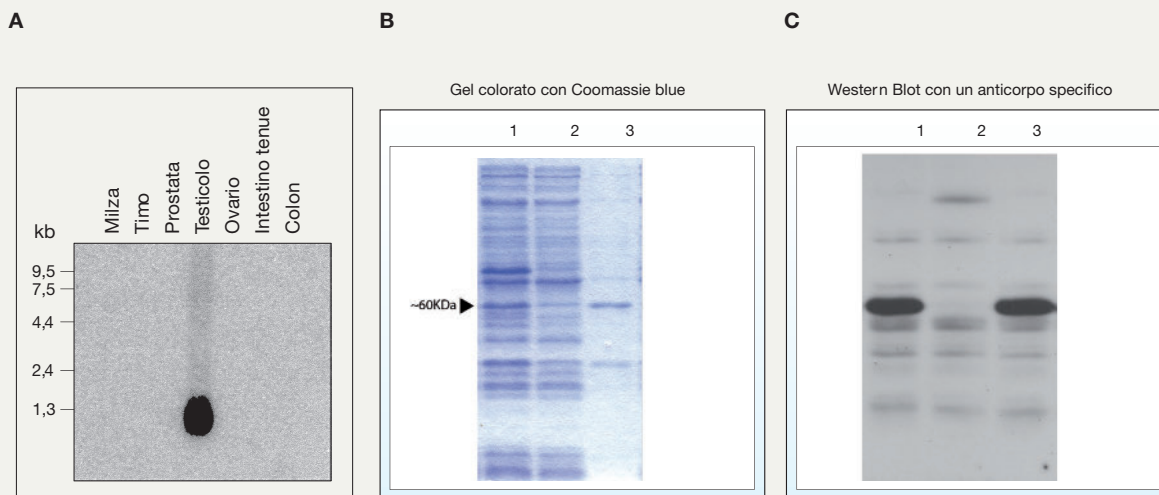


Figura 20F4.1 Northern blot e Western blot. (A) Identificazione di uno specifico mRNA in preparazioni di mRNA globali da diversi tessuti. Se un gene è espresso specificamente in un certo tessuto e non in altri, il corrispondente mRNA sarà presente solo in quel tessuto. Nel pannello è mostrato il risultato di un Northern blot eseguito con una sonda di DNA che dimostra che il gene in esame (che dà origine a un mRNA di circa 1,3 kb) è trascritto solo nel testicolo e non in altri tessuti. (B) Le proteine presenti in un estratto elettroforesi SDS-PAGE e colorate con Coomassie blue. Le proteine sono poi state purificate su una colonna cromatografica: le proteine che non si sono legate alla colonna sono state analizzate nella corsia 2, mentre le proteine legate ed eluite dalla colonna

sono state analizzate nella corsia 3. (C) Western blot delle stesse frazioni proteiche analizzate nel pannello (B) dopo trasferimento delle proteine su un filtro di nitrocellulosa e incubazione con un anticorpo che riconosce specificamente la proteina che si voleva purificare e che ha un peso molecolare di circa 60 kDa. Tramite l'analisi di Western blot si può notare che la proteina in esame si lega alla colonna cromatografica, mentre una proteina con circa lo stesso peso molecolare che non si lega alla colonna può essere evidenziata mediante colorazione con Coomassie blue, ma non viene rilevata dall'analisi Western. Le altre bande rilevate in tutte le corsie del Western blot sono proteine che interagiscono aspecificamente con gli anticorpi utilizzati.

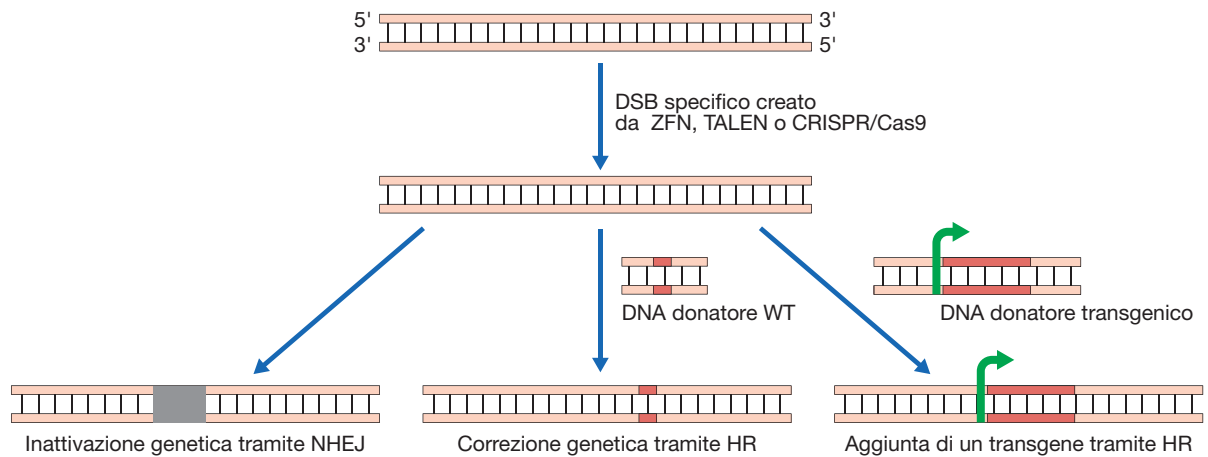


Figura 20.50 Modificazioni mirate nel genoma. L'utilizzo di nucleasi programmabili, quali le ZFN, TALEN e il sistema CRISPR/Cas9 permette di introdurre una rottura a doppio filamento sul DNA (DSB) in un punto specifico nel genoma. Questo taglio può essere riparato mediante NHEJ o mediante HR. La riparazione mediante NHEJ viene per lo più utilizzata per eliminare la funzio-

nalità del gene (knock-out), mentre la riparazione mediante HR può portare alla correzione di una mutazione esistente, oppure a introdurre nel genoma una specifica modificazione (ad es. un transgene) costruita in laboratorio dai ricercatori mediante l'applicazione di tecniche di ingegneria genetica.

reggere mutazioni esistenti o introdurre dei transgeni che, ad esempio, possono portare alla produzione di una proteina fluorescente (**fig. 20.50**).

Le procedure più efficienti per modificare il DNA in modo mirato sono tre: l'utilizzo di Zinc-Finger Nucleasi (ZFN), quello di TALE nucleasi (TALEN) e l'utilizzo della nucleasi batterica Cas9 (sistema CRISPR/Cas9).

Zinc-Finger e TALE nucleasi (ZFN e TALEN)

Le ZFN sono delle proteine ricombinanti formate da un dominio nucleasico che deriva da uno speciale enzima di restrizione, chiamato FokI, fuso a un dominio di legame al DNA che è costituito da almeno 3-4 “dita di zinco” (*vedi fig. 11.21*). È importante precisare che il dominio nucleasico di FokI può tagliare il DNA solo in forma di dimerico. Quindi, una ZFN è disegnata come una coppia di monomeri, ognuno contenente il dominio nucleasico di FokI fuso con 3-4 domini a dita di zinco che riconoscono sequenze specifiche diverse sui due filamenti di DNA e sono tra loro separate da una sequenza di 5-7 paia di basi (pb) che funziona da “spaziatore” (**fig. 20.51A**). Ciascuna “dita di zinco” è costituita da circa 30 amminoacidi e specifici residui nell' α -elica di ciascuna “dita di zinco” contattano 3 pb nel solco maggiore del DNA. La sequenza bersaglio sul DNA riconosciuta da una ZFN come dimerico è lunga 18-36 pb e, quindi, altamente specifica.

Le TALEN nucleasi sono simili alle ZFN: infatti, utilizzano ancora i due dimeri catalitici di FokI, ma un diverso e programmabile dominio di legame al DNA per indirizzare il taglio nucleasico di FokI sulla sequenza di DNA desiderata. TALE è l'acronimo di *Transcription Activator-Like Effector*: queste sono proteine secrete da batteri del genere *Xanthomonas* quando infettano una pianta. Il loro dominio di legame al DNA è costituito da 33-34 amminoacidi molto conservati, con l'eccezione delle posizioni 12 e 13 (chiamate *Repeat Variable Di-residues* o RVDs) che sono variabili e correlano con uno specifico nucleotide sul DNA. I moduli RVD Asn-Asn, Asn-Ile, His-Asp e Asn-Gly identificano, rispettivamente, G, A, C, T, e ciò permette di creare TALEN che, teoricamente, sono in grado di tagliare qualsiasi sequenza di DNA desiderata (**fig. 20.51B**).

Il lavoro sperimentale da fare per disegnare e costruire ZFN e TALEN utilizzabili per modificazioni mirate del genoma è, come si può immaginare, molto complesso. Infatti, dopo aver individuato la sequenza sul genoma che si vuole modificare attraverso la formazione di un DSB sul DNA, bisogna costruire le molecole di DNA in grado di codificare una specifica ZFN o TALEN e inserirle in un vettore opportuno mediante tecniche d'ingegneria genetica. Il vettore deve essere poi fatto entrare nella cellula bersaglio dove la sequenza di DNA corrispondente alla ZFN o TALEN specifica deve essere prima trascritta e tradotta, poi entrare nel nucleo della cellula e, infine, riconoscere e tagliare la sequenza di DNA bersaglio.

Il sistema CRISPR/Cas9

Il sistema CRISPR/Cas9 è stato originariamente individuato in molti batteri come un meccanismo di immunità che li protegge da infezioni da parte di batteriofagi, i virus dei batteri, e plasmidi. Questo tipo di immunità acquisita procede

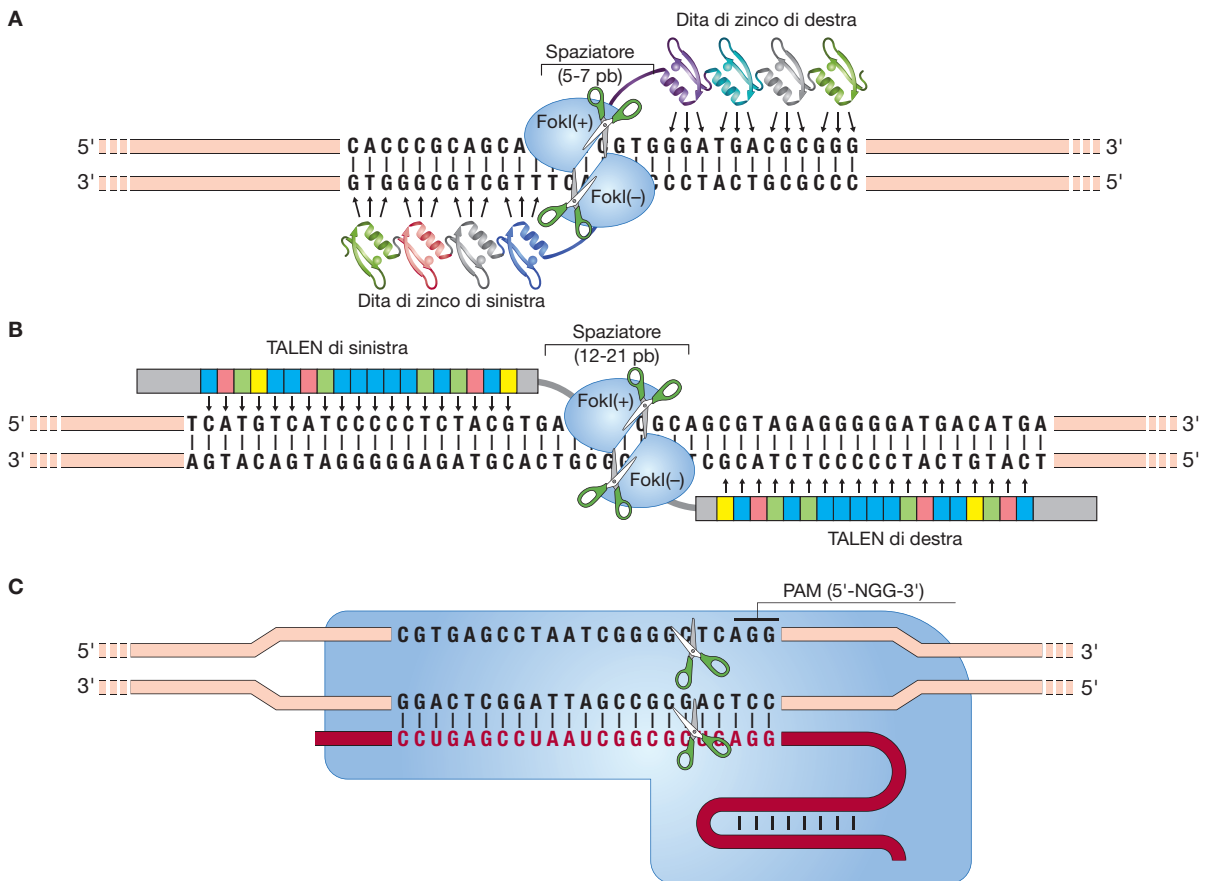


Figura 20.51 Meccanismo d'azione di nucleasi programmabili. (A) Zinc-Finger-Nucleasi (ZFN). In questa rappresentazione schematica è rappresentato un dimerico di ZFN legato al DNA. Ciascun monomero è composto dal dominio nucleasico di FokI (in azzurro) al C-terminale della proteina e, in questo specifico esempio, da 4 domini zinc-finger all'N-terminale della proteina. I 4 domini zinc-finger, attraverso interazioni specifiche con alcuni amminoacidi, riconosce 3 pb sul DNA. La sequenza bersaglio di una specifica ZFN è solitamente lunga 18-36 pb, escludendo lo spacer di 5-7 pb. Il sito di taglio sui due filamenti di DNA è schematizzato dalle forbici. (B) TALEN (*Transcription Activator-Like Effector Nucleasi*). Anche in questo pannello è rappresentato un dimerico di TALEN con

(in azzurro) il dominio nucleasico FokI al C-terminale. Al dominio N-terminale sono rappresentati, con dei rettangoli diversamente colorati, una serie di moduli TALE i cui colori diversi evidenziano come ciascun modulo identifichi uno specifico nucleotide sul DNA. La sequenza bersaglio di un dimerico TALEN è tipicamente lunga 30-40 pb, escludendo uno spacer di 12-21 pb. Il sito di taglio sui due filamenti di DNA è schematizzato dalle forbici. (C) Il sistema CRISPR/Cas9. La nucleasi Cas9 è indicata dal disegno colorato che è immaginario e non vuole suggerire alcuna struttura della nucleasi. In rosso è indicata la sequenza di un sgRNA (small guide RNA) che è complementare a 23 pb della sequenza di DNA bersaglio includendo il sito PAM (5'-NGG-3'), dove N, indica un qualsiasi nucleotide. Il sito di taglio sui due filamenti di DNA è schematizzato dalle forbici.

attraverso la cattura di frammenti di DNA di batteriofagi e plasmidi di circa 20 pb e l'incorporazione di questi frammenti di DNA estraneo (denominati protospacers) va a formare sul DNA batterico una regione denominata CRISPR, da *Clustered Regularly Interspaced Short Palindromic Repeats*. Tale biblioteca di DNA estraneo rappresenta la memoria delle infezioni subite e conferisce ai batteri un'immunità contro ulteriori infezioni. In tale regione, oltre alle sequenze di DNA derivate da batteriofagi e plasmidi, si trova anche la sequenza di DNA codificante per una nucleasi chiamata Cas9 (per *CRISPR associated*). Il sistema CRISPR/Cas9 agisce tagliando il DNA estraneo del batteriofago in un punto preciso che contiene uno specifico segnale per la nucleasi Cas9 (NGG), chiamato sequenza PAM (*Protospacer Adjacent Motif*). È stato anche dimostrato che, per tagliare il DNA nella sequenza PAM, la nucleasi Cas9 necessita di un piccolo RNA ausiliario, chiamato tracrRNA (*transactivating crRNA*). Durante un'infezione virale sono prodotti l'enzima Cas9, il piccolo tracrRNA oltre a un RNA per ciascuna sequenza virale registrata nel sito CRISPR (crRNA). Cas9 forma un complesso con il tracrRNA e con ciascun crRNA.

È stato successivamente dimostrato che è possibile riunire i due piccoli RNA indicati sopra in un singolo RNA "guida" (chiamato *small guide RNA* o sgRNA) per la nucleasi Cas9 (**fig. 20.51C**) e che il sgRNA è sufficiente per guidare Cas9 verso qualsiasi sequenza di DNA sintetizzata in provetta che contenga il motivo PAM (NGG). Queste caratteristiche conferiscono al sistema CRISPR/Cas9 una facilità e versatilità d'uso molto superiore rispetto alle ZFN e TALEN. Infatti, queste nucleasi sono concepite per produrre un DSB in funzione di una sequenza di DNA riconosciuta da una specifica proteina. Il sistema CRISPR/Cas9 supera questo limite perché il riconoscimento del sito di taglio sul DNA si basa su un frammento di RNA guida che è molto facile da costruire e Cas9 diventa solo lo strumento per eseguire il taglio nel punto previsto. L'utilizzo di CRISPR/Cas9 si è diffuso rapidamente anche perché è disponibile un software a libero accesso che permette di determinare il migliore RNA guida possibile. Inoltre i costrutti genetici per produrre questi RNA e la nucleasi Cas9 sono disponibili commercialmente. Questi sistemi necessitano ancora di ulteriori miglioramenti per escludere completamente la possibilità che possano avvenire tagli sul DNA al di fuori del sito previsto (effetti *off-targets*).

20.10 Valutazione dell'espressione di un gene

Una volta che un gene è stato clonato e magari sequenziato utilizzando l'una o l'altra delle tecniche descritte precedentemente (*vedi* parr. 20.4 e 20.8), uno degli interessi primari della Biologia Molecolare è conoscere tutti i dettagli molecolari che controllano l'espressione di quel gene. L'espressione del gene segue il flusso dell'informazione genetica: trascrizione, eventuale processamento dell'RNA e traduzione. Il livello globale del trascritto di un gene può essere misurato mediante esperimenti di **Northern blot** e il livello di proteina mediante **Western blot**, se si dispone di un anticorpo specifico diretto contro la proteina in esame (**finestra 20.4**).

È spesso di primaria importanza avere informazioni sulle regioni che controllano la trascrizione di un gene, cioè definire il promotore del gene stesso e altre regioni che ne controllano l'espressione. Il primo passo in tale direzione consiste nel definire il 5' del trascritto: cioè, stabilire a livello nucleotidico il punto d'inizio della trascrizione del gene. Esistono varie tecniche per ottenere tale informazione, ma la più utilizzata è chiamata **primer extension**.

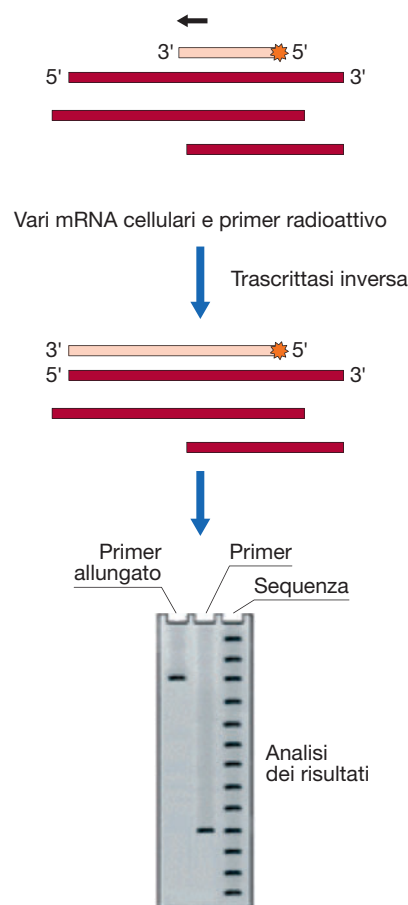
Se il gene in esame codifica per una proteina, si parte dalla purificazione degli mRNA da un tessuto o da un tipo cellulare in cui l'analisi Northern ha dimostrato che quel gene viene espresso a livello elevato. Alla miscela di tutti gli mRNA è aggiunto un oligonucleotide (primer) radioattivo complementare a una regio-

ne vicina all'inizio della regione codificante (ORF) del gene di cui, ovviamente, deve essere nota la sequenza. Si forma quindi un ibrido RNA/DNA tra l'mRNA del gene e il DNA primer complementare: il 3' del primer può essere allungato fino a raggiungere l'estremità 5' dell'mRNA se nella reazione vengono aggiunti i 4 dNTP e la trascrittasi inversa, l'enzima in grado di copiare RNA in DNA. La posizione del 5' del trascritto viene quindi determinata misurando di quanto è stato allungato il primer radioattivo e correndo sul gel in corsie parallele una sequenza di DNA effettuata utilizzando lo stesso primer (**fig. 20.52**).

Stabilito il 5' di un trascritto, è importante definire con precisione il promotore e le regioni di controllo del gene in esame. Questo tipo di analisi è molto spesso impossibile o comunque difficile da eseguire analizzando il promotore nella sua localizzazione genomica naturale. Per aggirare questo problema si fa spesso uso dei "geni reporter".

I **geni reporter** codificano, per lo più, per proteine di cui è facile misurare la quantità e l'attività. I reporter più utilizzati includono i geni codificanti per la **β -galattosidasi** (una proteina batterica che, abbiamo già visto, può essere facilmente rivelata utilizzando un substrato cromogeno o in grado di generare fluorescenza), la **proteina fluorescente verde** o **GFP** (una proteina fluorescente naturale che verrà descritta ampiamente più avanti) e la **luciferasi** (la proteina responsabile della luminescenza delle lucciole, la cui attività enzimatica è misurata mediante l'emissione di luce visibile). Con tecniche di DNA ricombinante classiche, a monte della sequenza codificante per uno o per l'altro di questi geni reporter è introdotta la regione che si presuppone possa contenere il promotore e

Figura 20.52 Saggio di estensione del primer per definire il 5' di un trascritto.



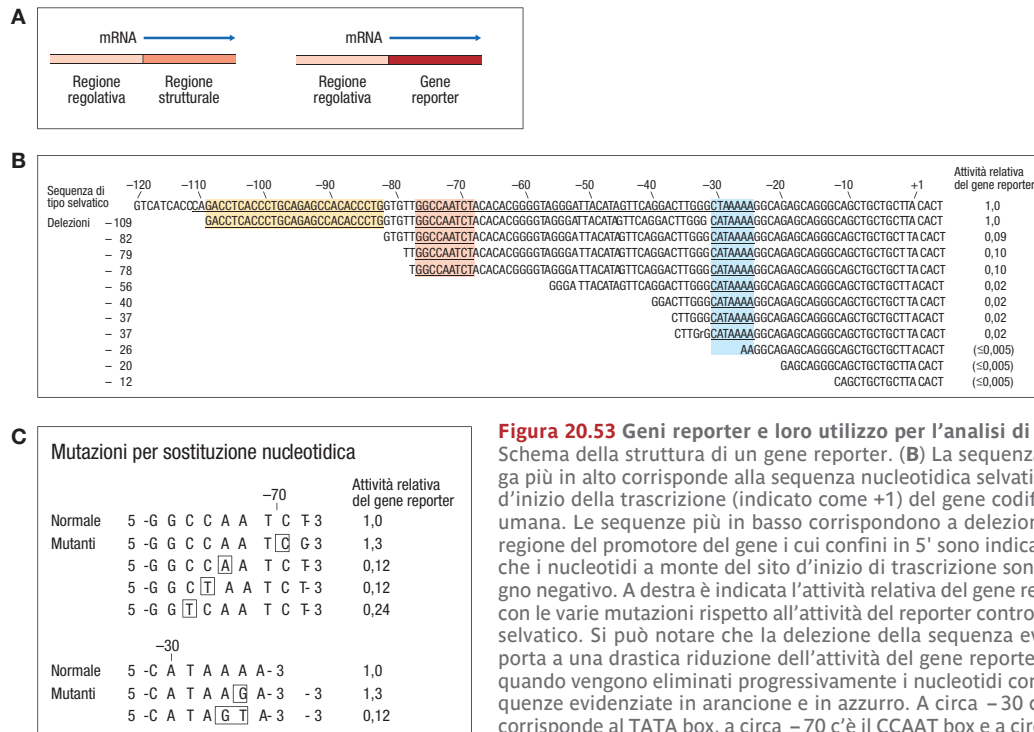


Figura 20.53 Geni reporter e loro utilizzo per l'analisi di un promotore. (A) Schema della struttura di un gene reporter. **(B)** La sequenza mostrata nella riga più in alto corrisponde alla sequenza nucleotidica selvatica a monte del sito d'inizio della trascrizione (indicato come +1) del gene codificante la β -globina umana. Le sequenze più in basso corrispondono a delezioni progressive della regione del promotore del gene i cui confini in 5' sono indicati a sinistra. Si noti che i nucleotidi a monte del sito d'inizio di trascrizione sono indicati con il segno negativo. A destra è indicata l'attività relativa del gene reporter nei costrutti con le varie mutazioni rispetto all'attività del reporter controllato dal promotore selvatico. Si può notare che la delezione della sequenza evidenziata in giallo porta a una drastica riduzione dell'attività del gene reporter. Lo stesso accade quando vengono eliminati progressivamente i nucleotidi corrispondenti alle sequenze evidenziate in arancione e in azzurro. A circa -30 c'è la sequenza che corrisponde al TATA box, a circa -70 c'è il CCAAT box e a circa -100 è presente in duplice copia il motivo CACCC. **(C)** È qui evidenziata in dettaglio l'importanza del CCAAT box a circa -70 e del TATA box a circa -30; infatti, mutazioni per sostituzione di singoli nucleotidi causano drastici effetti sull'attività del gene reporter.

le regioni regolatrici del gene che vogliamo analizzare. Ne deriva che l'espressione del gene reporter è ora regolata da questo promotore. Il costrutto in questione è introdotto all'interno delle cellule e, per esempio, si può valutare l'espressione del gene reporter in svariate condizioni sperimentali (ad es. a diversi stadi dello sviluppo, durante il ciclo cellulare ecc.). I geni reporter sono poi utilizzati per definire l'"anatomia molecolare" del promotore e di altre regioni di controllo. Infatti, attraverso delezioni progressive si possono determinare i confini in 5' e in 3' del promotore e attraverso la mutagenesi sito-specifica (**finestra 20.5**) si può valutare la rilevanza funzionale della sequenza del promotore a livello dei singoli nucleotidi (**fig. 20.53**).

20.11 Inibizione dell'espressione di geni specifici: RNA antisenso e interferenza da RNA

Uno dei modi per capire come funziona una macchina complicata come una cellula è di vedere cosa succede se si elimina un pezzo alla volta; in particolare, cosa succede inibendo l'espressione di singoli geni. Tale approccio sperimentale viene anche chiamato **Genomica funzionale** o **Genetica inversa** (**finestra 20.5**). L'eliminazione di un gene si ottiene con tecniche di knock-out genetico che variano a seconda degli organismi utilizzati nell'esperimento e che saranno, pertanto, descritte nel capitolo 22, in cui sono presentati i più importanti organismi modello utilizzati in Biologia Molecolare.

Per spegnere l'espressione di un gene era stata messa a punto, fin dagli anni ottanta del secolo scorso, la tecnica dell'RNA antisenso, cioè l'introduzione nella cellula di molecole di un RNA che, essendo complementare a un mRNA bersaglio, vi

FINESTRA 20.5

MUTAGENESI SITO-SPECIFICA O MUTAGENESI MIRATA

Per quasi un secolo la Genetica classica si è avvalsa, per studiare organismi procariotici ed eucariotici, di mutazioni che si trovano in natura o che si possono ottenere in laboratorio aumentando la frequenza di mutazione mediante mutagenesi chimica o radiazioni; in ogni caso, si trattava di mutazioni casuali. In questi ultimi quarant'anni i progressi a livello molecolare ci hanno dato, invece, la possibilità di produrre *in vitro* mutazioni in maniera mirata; cioè, la possibilità di cambiare a piacimento uno o alcuni nucleotidi di una sequenza che poi può eventualmente essere reintrodotta *in vivo* per analizzarne l'effetto. Questo processo viene talvolta indicato con il nome di **Genetica inversa**: infatti, anziché partire dal fenotipo per arrivare al genotipo come fa la Genetica classica, si parte dalla sequenza di un gene, per inattivarlo completamente o per introdurre le mutazioni desiderate e se ne va poi a vedere l'effetto a livello fenotipico.

Sono state sviluppate varie procedure per la mutagenesi sito-specifica. La figura mostra la procedura di base relativamente semplice per produrre una mutazione mirata di un singolo

nucleotide in un gene o tratto di DNA. Questo deve, innanzitutto, essere clonato in un vettore a singolo filamento come, tipicamente, il fago M13. Il DNA a singolo filamento purificato viene associato a un oligonucleotide di 20-30 nt complementare alla regione di interesse, tranne che per il nucleotide che si vuole mutare, al centro della sequenza. Anche se non perfettamente complementare, l'oligonucleotide si appaierà con la sequenza genica e potrà funzionare da innesco per una reazione di allungamento portata avanti dalla DNA polimerasi aggiunta alla soluzione. Le molecole di DNA a doppia elica che si formano, costituite da un filamento normale e uno mutato, vengono introdotte in *E. coli* mediante trasformazione. Le molecole di DNA andranno incontro a cicli di replicazione che porteranno alla formazione di alcuni genomi a doppia elica normali e di altri con la mutazione, che andranno a distribuirsi nella progenie batterica. I batteri che portano la sequenza mutata possono poi essere selezionati mediante la tecnica di ibridazione su colonia utilizzando come sonda lo stesso oligonucleotide usato per la mutagenesi, reso in qualche maniera radioattivo.

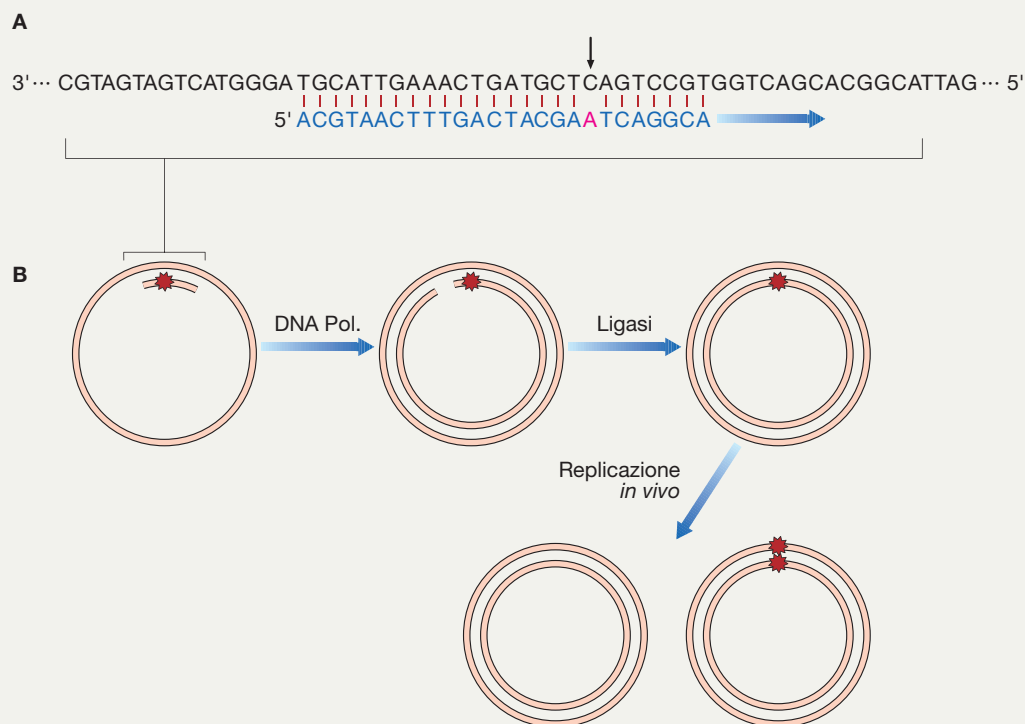
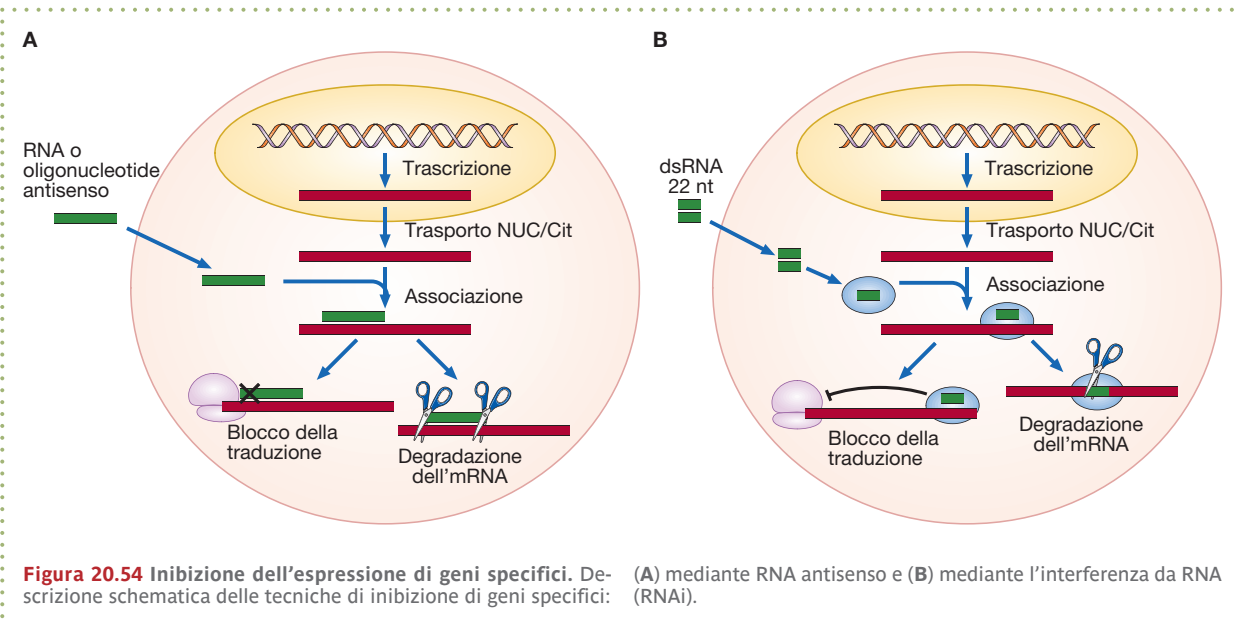


Figura 20F5.1 Procedura di base per la mutagenesi sito-specifica. Il frammento di DNA in cui si vuole introdurre una mutazione puntiforme è clonato in un vettore virale a singolo filamento (fago M13). (A) Visione dettagliata della sequenza della regione in cui si vuole introdurre la mutazione C → T (indicata con una freccia). A questa regione viene ibridato un oligonucleotide (blu) perfettamente complementare al filamento originale, tranne che nella posizione che deve essere mutata, che contiene una A (rosso) anziché una G.

Questo oligonucleotide funziona da innesco per la sintesi di DNA da parte della DNA polimerasi che allunga l'estremità 3'. (B) Visione globale della molecola circolare a singolo filamento in cui si vuole introdurre la mutazione. L'oligonucleotide contenente la mutazione (asterisco rosso) viene allungato dalla DNA polimerasi, e il filamento di nuova sintesi chiuso dalla ligasi. Il DNA viene introdotto in *E. coli*, dove mediante replicazione semiconservativa produrrà 50% di molecole a doppia elica mutate e 50% di tipo selvatico.



si appaia e ne inibisce la traduzione (**fig. 20.54A**). Gli RNA antisenso sintetici, pur essendosi dimostrati utili in alcuni casi, presentavano però vari problemi relativi al loro trasferimento nella cellula, a degradabilità, selettività del bersaglio, citotossicità ed effetti collaterali. La situazione è stata migliorata utilizzando, come molecole antisenso, oligonucleotidi fosforotioati (S-DNA) od oligonucleotidi morfolino, che presentano una maggiore stabilità. Oggi, però, tale tecnologia è stata quasi totalmente soppiantata dall'**interferenza da RNA (RNAi, RNA interference)**. Questa tecnica è stata messa a punto abbastanza recentemente ed è oggi molto utilizzata: sfruttando il principio di funzionamento dei microRNA naturali (*vedi* capp. 12, 16 e 18), permette di inibire *in vivo* l'espressione di geni specifici. L'osservazione che ha permesso di sviluppare questo metodo è stata che l'introduzione, mediante trasfezione o microiniezione in cellule di vari organismi quali *C. elegans* e *Drosophila*, di molecole di RNA a doppio filamento, uno dei quali, complementare a un mRNA cellulare, porta alla degradazione di quest'ultimo. Ciò avviene perché l'RNA a doppio filamento viene tagliato in brevi frammenti di 22 nt di lunghezza (chiamati spesso **siRNA** o **small interfering RNA**), maturato e assemblato in complessi RISC che poi vanno a interferire con la stabilità e/o traduzione dell'mRNA bersaglio. Il significato fisiologico di tutto questo è che, per una cellula, RNA estraneo a doppio filamento vuol dire in genere infezione virale o retrotrasposizione, ambedue fenomeni da cui la cellula deve difendersi. Quindi, introducendo nella cellula molecole di RNA a doppio filamento con sequenza complementare a un mRNA cellulare si ottiene la degradazione specifica di quest'ultimo.

Tutto ciò funziona molto bene in alcuni organismi eucariotici modello (*vedi* cap. 22), ma quando si è andati a valutare la tecnica in cellule umane o di altri mammiferi non si riuscivano a ottenere risultati significativi. Si è presto capito che il motivo sta nel fatto che nei mammiferi esistono almeno tre diversi meccanismi per la difesa dall'invasione da molecole di RNA a doppia elica. Infatti, oltre al sistema dei microRNA (che è quello che ci interessa perché agisce su bersagli specifici) ci sono almeno altre due vie di difesa: una, mediata dalla RNasi L, che causa la degradazione aspecifica di tutto l'mRNA; l'altra, mediata dalla chinasi PKR, che, fosforilando il fattore d'inizio della traduzione eIF2, porta al blocco dell'inizio di tutta la sintesi proteica. In queste condizioni la cellula muore e non siamo in grado di studiare l'effetto dell'inibizione specifica dell'mRNA bersaglio. Si è, però, trovato che questi ultimi due meccanismi funzionano solo in risposta

a RNA a doppia elica abbastanza lunghi, comunque più lunghi di 26 pb. Invece il meccanismo di RNA interference viene innescato anche da RNA a doppia elica più corti. Pertanto, la tecnica oggi usata per inibire geni specifici anche in cellule umane si basa sulla trasfezione nelle cellule di RNA di 22 nt a doppia elica, una delle quali è complementare a una regione dell'mRNA bersaglio che si vuole inibire (**fig. 20.54B**). In queste condizioni, senza bisogno di previa maturazione, i due filamenti separati di 22 nt vengono assemblati in complessi RISC (*vedi* par. 18.2). Abbiamo così inibizione dell'mRNA bersaglio, ma non degradazione aspecifica di tutti gli mRNA e nemmeno blocco della sintesi proteica totale. La tecnica è efficiente e relativamente semplice: oggi viene usata in tutti i laboratori di Biologia Molecolare per inibire specificamente l'espressione di uno o più geni. Essendo basata sullo spegnimento dell'espressione di un gene, tale tecnica è anche indicata come **knock-down di un gene**. Lo svantaggio di tale metodologia è che lo spegnimento dell'espressione del gene in esame (misurabile mediante RT-PCR o **Western blotting** della proteina codificata dall'mRNA in questione) può essere totale o solo parziale a seconda dei casi. Inoltre, non è possibile escludere totalmente che l'effetto del siRNA verso uno specifico mRNA non abbia anche un qualche effetto collaterale non prevedibile sull'espressione di altri geni. L'inattivazione completa di uno specifico gene è invece assicurata da esperimenti di distruzione genica o knock-out come quelli che saranno descritti nel capitolo 22.

20.12 Interazioni tra macromolecole biologiche

Uno degli obiettivi principali della Biologia Molecolare consiste nello studio della struttura e funzione di macromolecole biologiche. Le diverse macromolecole all'interno della cellula (quali, per esempio, acidi nucleici e proteine) svolgono funzioni diverse che sono, però, funzionalmente collegate tra di loro. Infatti, DNA e RNA interagiscono fisicamente con numerose proteine che hanno il ruolo di modulare e regolare l'espressione dei geni. Talvolta, i componenti sono così numerosi da formare vere e proprie strutture nucleoproteiche, quali nucleosomi, ribosomi, spliceosomi ecc. Come abbiamo anticipato diverse volte nel testo, numerose proteine interagiscono fisicamente con acidi nucleici, ma possono anche interagire tra di loro, al punto che la maggior parte delle funzioni di proteine coinvolte in vari aspetti del metabolismo cellulare sono portate avanti non da proteine che agiscono singolarmente, ma da vere e proprie "macchine" proteiche composte da numerosi elementi. Qui di seguito descriveremo alcune delle tecnologie più utilizzate per lo studio delle interazioni proteina-DNA e proteina-proteina, lasciando ad altri testi una trattazione di metodologie più complesse.

Interazioni proteina-DNA

Numerose proteine interagiscono con gli acidi nucleici, sia DNA che RNA. Tali interazioni influenzano essenzialmente tutti gli aspetti del metabolismo degli acidi nucleici: replicazione, riparazione, ricombinazione e trascrizione del DNA, processamento e stabilità dell'RNA e traduzione di molecole di mRNA.

Si hanno oggi informazioni strutturali precise su circa 400 complessi DNA-proteina e 100 complessi RNA-proteina. I dettagli strutturali e chimico-fisici che mediano tali interazioni vanno al di là degli obiettivi di questo testo di Biologia Molecolare di base. Basti qui ricordare che le interazioni tra proteine e acidi nucleici coinvolgono diversi tipi di legami chimici, quali i legami idrogeno, legami idrofobici, forze di van der Waals e interazioni elettrostatiche. Le interazioni tra proteine e acidi nucleici coinvolgono preferenzialmente catene laterali di particolari amminoacidi (arginina, lisina, serina, asparagina e glutammina sono i più frequenti) e particolari atomi di specifiche basi azotate (nel DNA la più coinvolta è la guanina). Esistono, però, delle differenze tra i tipi di legami e le strutture

coinvolte nel modulare le interazioni proteina-DNA e proteina-RNA. Per esempio, nelle interazioni tra proteine e DNA a doppio filamento sono più frequenti le interazioni che coinvolgono legami idrogeno rispetto a forze di van der Waals; tale rapporto è invertito nelle interazioni proteina-RNA. Inoltre, le interazioni proteina-DNA a doppio filamento coinvolgono frequentemente legami con l'impalcatura zucchero-fosfato. Tale tipo d'interazioni è molto più raro nell'RNA, probabilmente perché l'RNA è a singolo filamento e molte basi azotate non si appaiano tra loro. È altresì noto che le molecole di RNA formano strutture secondarie complesse e queste, a loro volta, possono influenzare le interazioni con specifiche proteine.

Le interazioni tra proteine e acidi nucleici possono essere altamente specifiche, coinvolgendo particolari sequenze nucleotidiche o strutture sugli acidi nucleici e caratteristici domini per quanto riguarda le proteine. Alternativamente, soprattutto per le interazioni tra proteine basiche e l'impalcatura zucchero-fosfato del DNA a doppio filamento, le interazioni proteina-DNA possono essere guidate quasi esclusivamente da interazioni elettrostatiche (ad es. le interazioni tra DNA e istoni nella formazione dei nucleosomi sulla cromatina).

Le tecniche utilizzate per studiare le interazioni proteine-acidi nucleici sono in gran parte sovrapposte sia per quanto riguarda il DNA che l'RNA, per cui ci limiteremo a descrivere le più semplici e comuni nell'analisi di un'interazione tra una specifica proteina e il DNA.

■ Saggio di “gel-shift” o EMSA (*Electrophoretic Mobility Shift Assay*)

Questo saggio è comunemente utilizzato per verificare se esistono una o più proteine che legano specificamente una certa sequenza di DNA o, in generale, un segmento di DNA che contiene una determinata sequenza.

La procedura è semplice e può venire utilizzata per verificare se in un estratto cellulare esiste una proteina in grado di legare un certo segmento di DNA, e lo stesso saggio può poi essere utilizzato per purificare la proteina in esame. La tecnica si basa sul concetto che se una proteina si lega a un segmento di DNA, la massa del segmento aumenta provocando un ritardo nella sua migrazione elettroforetica (**finestra 20.3**). Dato che i quantitativi di DNA utilizzati nel saggio sono limitati, è difficile visualizzarlo mediante colorazione con bromuro d'etidio. Si utilizza, perciò, un frammento di DNA marcato all'estremità 5' con un atomo di fosforo radioattivo e si evidenziano la posizione del frammento libero o ritardato dopo autoradiografia (**fig. 20.55**).

Quando il saggio EMSA è eseguito utilizzando estratti cellulari non purificati, si possono trovare proteine che legano il segmento di DNA in modo aspecifico mediante, per esempio, semplici interazioni elettrostatiche. Legami proteina-DNA specifici possono essere distinti da legami aspecifici mediante esperimenti di competizione in cui si usa, come competitore per il legame, un eccesso dello stesso segmento di DNA non radioattivo. Se tale competitore è presente in quantità 100 volte più alta dello stesso DNA radioattivo, la proteina legherà tale DNA competitore con una probabilità 100 volte più alta rispetto al DNA radioattivo, facendo essenzialmente scomparire il segnale corrispondente al DNA ritardato. Se l'interazione è mediata dalla specifica sequenza nucleotidica indicata in **figura 20.55**, un DNA contenente anche soltanto una mutazione puntiforme in tale sequenza non sarà più in grado di funzionare come un efficiente competitore.

■ Saggio di DNA “footprinting”

Il saggio di DNA “footprinting” permette di definire con precisione la regione di DNA con cui una certa proteina prende contatto. La tecnica richiede il preventivo isolamento di un segmento di DNA con cui tale proteina interagisce.

Il saggio si basa sul fatto che una proteina che lega il DNA lo protegge, nella regione di contatto, dall'attacco da parte di agenti chimici o di nucleasi che ta-

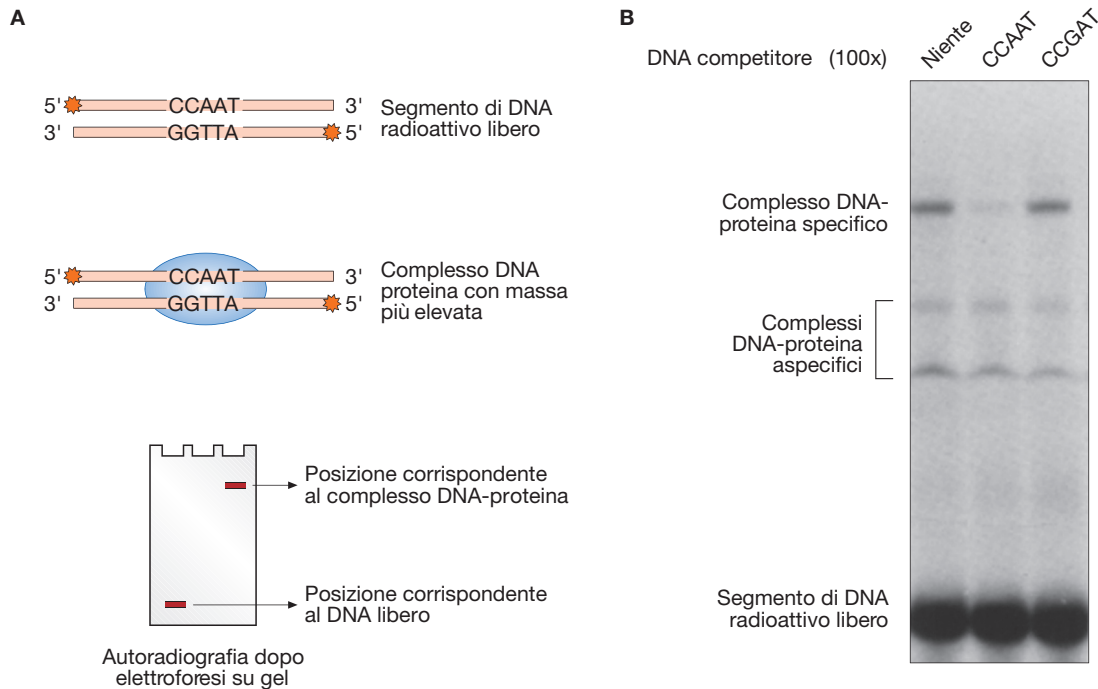


Figura 20.55 Saggio di “gel-shift” o EMSA (*Electrophoretic Mobility Shift Assay*). (A) Schema riassuntivo della funzionalità del saggio EMSA (vedi testo); gli asterischi arancione alle estremità 5' indicano la marcatura con ^{32}P (fosforo radioattivo). (B) Il saggio è stato eseguito su un estratto grezzo di embrioni di *Xenopus* utiliz-

zando come sonda radioattiva un frammento di DNA di circa 40 nt corrispondente a un tratto del promotore del gene *GATA2* che contiene la sequenza CCAAT a cui si lega una proteina chiamata CBTF. Il legame specifico tra CBTF e il frammento di DNA è dimostrato da esperimenti di competizione come descritto nel testo.

gliano aspecificamente il DNA (quale la DNasi I). Il frammento di DNA viene marcato a una delle due estremità 5' mediante l'attacco di un atomo di fosforo radioattivo. Si procede poi a una digestione parziale con la DNasi I, cosicché, mediamente, ogni molecola di DNA venga tagliata una sola volta, in una qualsiasi delle posizioni nucleotidiche del frammento. Si genera così una serie di frammenti radioattivi la cui lunghezza differisce per un singolo nucleotide. Mancheranno, però, in tale serie di frammenti, quelli corrispondenti alle zone protette dalla proteina, in quanto la DNasi I non riuscirà a tagliare in tali regioni (**fig. 20.56**). È possibile marcare una o l'altra delle due estremità 5' del frammento: in tal modo, l'esperimento eseguito in duplicato, permette di determinare la zona protetta dalla proteina su entrambi i filamenti di DNA. I confini delle regioni protette sono determinati paragonando la serie di bande radioattive ottenute dopo digestione con DNasi I in presenza o in assenza della proteina che lega il DNA. Le bande corrispondenti ai frammenti radioattivi sono evidenziate mediante autoradiografia dopo elettroforesi in condizioni denaturanti (per separare i due filamenti di DNA).

■ SELEX (*Systematic Evolution of Ligands by Exponential enrichment*)

La grande versatilità dell'RNA nel formare strutture diverse e complesse, con capacità di legare molecole o avere attività catalitica, ha suggerito la messa a punto di una tecnica di “evoluzione *in vitro*” che ha permesso di isolare RNA con varie proprietà quali, ad esempio, RNA che agiscono come veri e propri enzimi (RNA ligasi o RNA replicasi) o che hanno la capacità di legare antibiotici o altri substrati d'interesse. Analogamente la tecnica può essere applicata a molecole di DNA per identificare sequenze che legano bersagli specifici.

In particolare, Ellington e Szostak a Harvard e Tuerk e Gold a Boulder, Università del Colorado, hanno sviluppato una nuova tecnica per identificare mole-

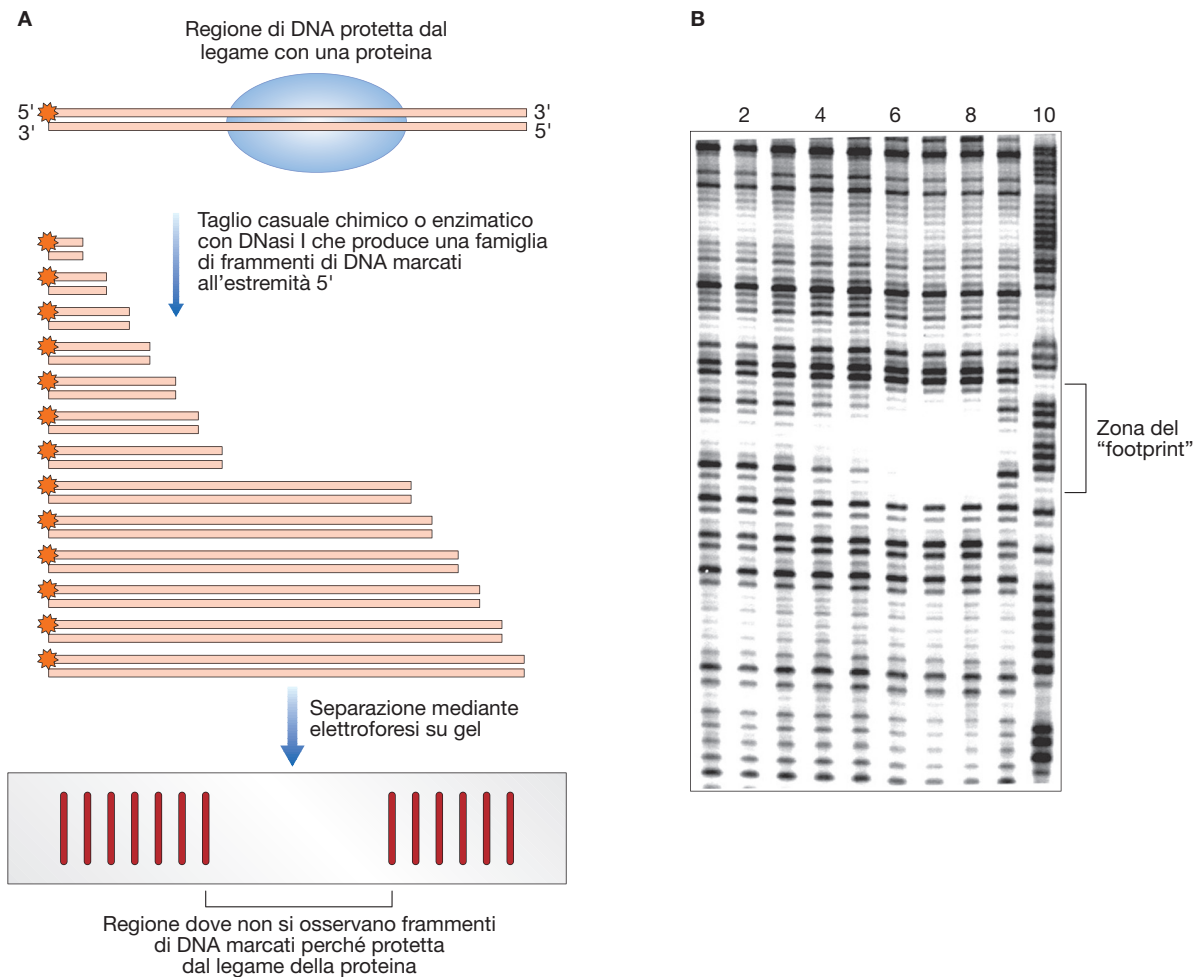


Figura 20.56 Saggio di DNA “footprinting”. (A) Schema riassuntivo della procedura. L’asterisco all’estremità 5' di uno dei due filamenti del frammento di DNA indica un atomo di fosforo radioattivo. Le bande radioattive differiscono una dall'altra per un singolo nucleotide. (B) Esempio di un esperimento di DNA footprinting in cui diverse frazioni di una colonna cromatografica per

la purificazione della proteina che si lega al DNA sono state analizzate nelle diverse corsie. La proteina che lega il DNA si trova nelle frazioni 6-8. La posizione dei nucleotidi corrispondenti ai confini della zona di “footprinting” protetta dal legame con la proteina dal taglio con DNasi I può essere definita con estrema precisione.

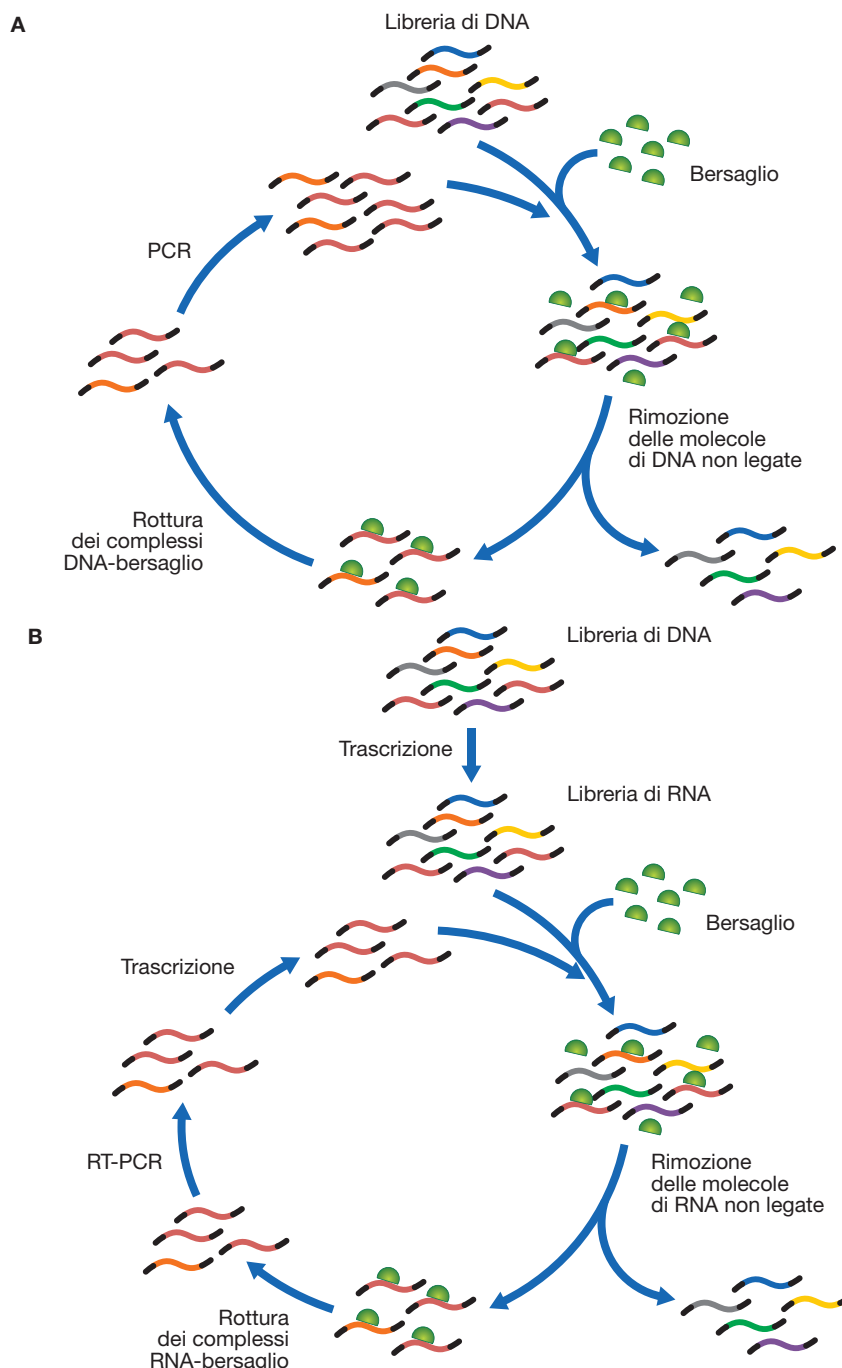
cole di RNA in grado di interagire specificamente con piccole molecole, definite aptameri, dal latino *aptus* (adatto) e dal greco *meros* (particella) (vedi fig. 2.50).

Questa tecnica, denominata SELEX (*Systematic Evolution of Ligands by Exponential enrichment*), permette di isolare da una libreria di miliardi di molecole di RNA o DNA, la particolare molecola che porta avanti la funzione per la quale è stata selezionata (fig. 20.57). La libreria è costituita da molecole di DNA o RNA che contengono nella parte centrale sequenze nucleotidiche casuali (randomizzate) e alle estremità 5' e 3' tratti a sequenza nota per permettere il legame di primers per una successiva amplificazione delle molecole selezionate tramite PCR.

La libreria costituita, come già accennato, da un pool di miliardi di molecole diverse, viene fatta interagire con un substrato non-specifico per eliminare il rumore di fondo e, solo successivamente, è fatta interagire con il bersaglio specifico prescelto, che può essere una proteina, un antibiotico come nell'esempio fatto nella figura 2.50, o qualsiasi molecola di interesse. Un certo numero di molecole presenti nella libreria potranno legarsi al bersaglio che viene isolato insieme alle molecole di RNA/DNA che con esso hanno interagito. Dopo vari lavaggi per

Figura 20.57 Illustrazione schematica della metodologia SELEX.

Una libreria di molecole diverse di DNA (A) o RNA (B) viene fatta interagire con un bersaglio specifico, che può essere una proteina che si lega al DNA (o all'RNA) o qualsiasi molecola di interesse. La libreria di molecole contiene nella parte centrale sequenze casuali (con colori diversi) e alle estremità 5' e 3' (disegnati in nero) tratti a sequenza nota necessari per legare i primers richiesti per successive amplificazioni mediante PCR. Dopo vari cicli di selezione e amplificazione (*vedi testo*) si isolano le molecole che hanno maggiore affinità con il bersaglio (molecole rosse nella figura).



eliminare interattori non-specifici o che si sono legati molto debolmente, gli interattori specifici sono eluiti e amplificati tramite PCR come indicato sopra e mostrato nella **figura 20.57A**. Il ciclo sarà ripetuto più volte per arricchire la popolazione di molecole che legano specificamente il bersaglio. Nel caso di aptameri di RNA sarà necessario un passaggio tramite “trascrizione inversa” (**fig. 20.57B**). La fase finale prevede il sequenziamento e clonaggio della molecola di interesse così individuata.

Questa tecnica si è rivelata molto potente per disegnare e individuare aptameri che hanno una funzione in medicina e/o isolare veri e propri enzimi a RNA.

■ Saggio di immunoprecipitazione della cromatina o ChIP (*Chromatin Immuno Precipitation*)

Il saggio di DNA footprinting è normalmente eseguito *in vitro*; tuttavia, con opportune modifiche può essere eseguito anche *in vivo*. Per la sua sensibilità, però, lo studio delle interazioni proteina-DNA *in vivo* è oggi preferenzialmente eseguito con il saggio di immunoprecipitazione della cromatina o ChIP.

In questo saggio, le cellule vengono trattate con formaldeide, un reagente in grado di formare “cross-link” che legano covalentemente proteine e DNA che interagiscono tra di loro (**fig. 20.58**). Dopo la lisi delle cellule, il DNA viene frammentato in segmenti di 300-500 pb mediante sonicazione. I frammenti di DNA che legano *in vivo* una specifica proteina possono venire separati da tutti gli altri frammenti di DNA disponendo di un anticorpo specifico contro la proteina in esame. I complessi antigene-anticorpo con attaccati i frammenti di DNA con cui la proteina interagisce sono, infatti, immunoprecipitati e purificati utilizzando una resina capace di legare specificamente gli anticorpi, quale, per esempio, una resina contenente la proteina A, una proteina batterica che interagisce fortemente con una regione presente in tutte le immunoglobuline. Riscaldando a 65 °C i campioni purificati tramite immunoprecipitazione è possibile far revertire il cross-linking tra i frammenti di DNA e la proteina in esame, così da poter procedere alla purificazione dei frammenti di DNA stessi e alla loro amplificazione tramite PCR ed eventuale sequenziamento. È atteso che il frammento con cui la proteina ha interagito sia arricchito in un campione incubato con l'anticorpo specifico contro la proteina in esame, rispetto a un campione trattato con un anticorpo di controllo.

■ Risonanza plasmonica superficiale (SPR, *Surface Plasmon Resonance*)

Interazioni tra DNA e proteine (ma anche tra proteina-proteina e, in genere, tra macromolecole biologiche) possono essere analizzate con un biosensore in grado

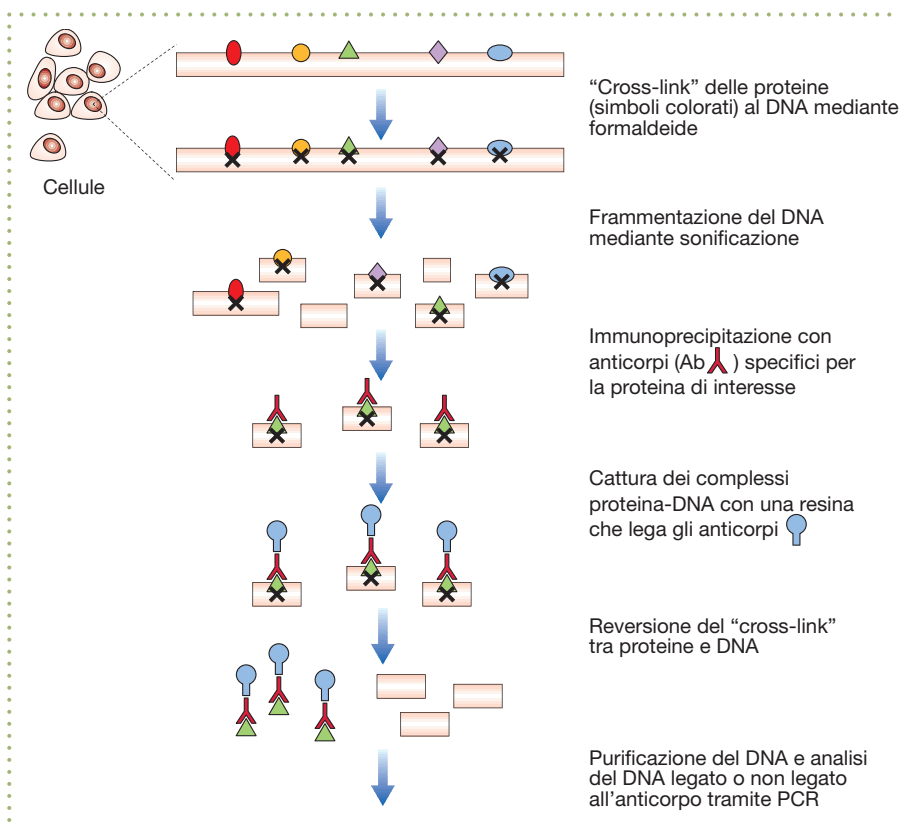


Figura 20.58 Saggio di immunoprecipitazione della cromatina (ChIP).

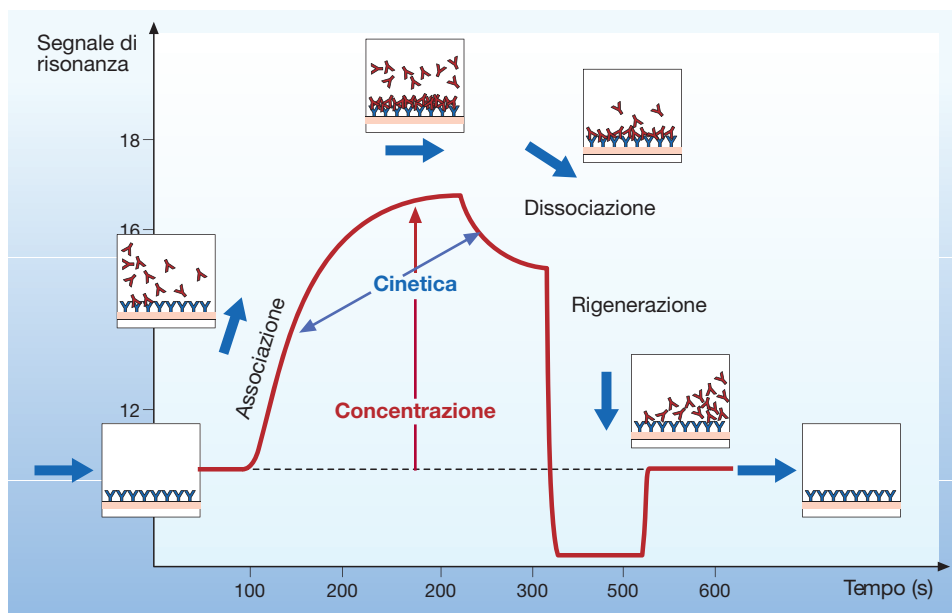


Figura 20.59 Saggio di risonanza plasmatica superficiale (SPR). Le modalità del saggio sono descritte nel testo. Le due macromolecole in esame sono indicate dai simboli rossi e blu. La macromolecola con il simbolo blu (nell'esempio, DNA) è quella che viene

legata alla superficie metallica (rosa), mentre la macromolecola con il simbolo rosso (nell'esempio, proteina) è quella che si trova in soluzione. Durante il passaggio di rigenerazione del biosensore, la molecola interagente (rossa) viene eliminata.

di trasformare l'interazione fisica tra due macromolecole in un segnale che possa essere facilmente misurato. Senza entrare nei dettagli del biosensore SPR, il suo funzionamento può essere così semplificato: una delle due macromolecole di cui si vuole analizzare l'interazione (ad es. DNA) viene immobilizzata su una superficie metallica, mentre la seconda macromolecola (ad es. una proteina in soluzione) viene fatta passare sulla superficie metallica, così che si possa formare un complesso DNA-proteina nel caso in cui le due macromolecole interagiscano tra di loro. La formazione di tale complesso modifica l'indice di rifrazione della soluzione che passa sul film metallico quando questa è colpita da un raggio di luce polarizzata di una determinata lunghezza d'onda. Come mostrato in **figura 20.59**, questa tecnica permette di ottenere anche indicazioni sulle cinetiche di associazione e dissociazione delle due molecole in esame. Il saggio è relativamente semplice, sensibile e richiede pochi passaggi sperimentali, ma l'acquisto del biosensore è piuttosto costoso.

■ Microscopia a forza atomica (AFM, *Atomic Force Microscopy*)

La microscopia a forza atomica sta diventando una delle tecniche di microscopia più utilizzate per ottenere informazioni su interazioni proteina-DNA; tali interazioni possono essere evidenziate anche tramite microscopia elettronica a scansione. Il principale vantaggio dell'AFM rispetto alla microscopia elettronica consiste nel fatto che le immagini di AFM offrono un profilo tridimensionale e non bidimensionale del complesso proteina-DNA e la preparazione dei campioni non richiede alcun trattamento speciale.

Il microscopio a forza atomica consiste in una microleva alla cui estremità è montata una punta acuminata, tipicamente di silicio, che ha un raggio di curvatura dell'ordine dei nanometri. La punta viene collocata nelle strette vicinanze della superficie del campione di cui si vuole effettuare la scansione (**fig. 20.60**). L'interazione tra la punta del microscopio e il campione provoca una deflessione della microleva che viene misurata utilizzando un raggio laser riflesso dalla som-

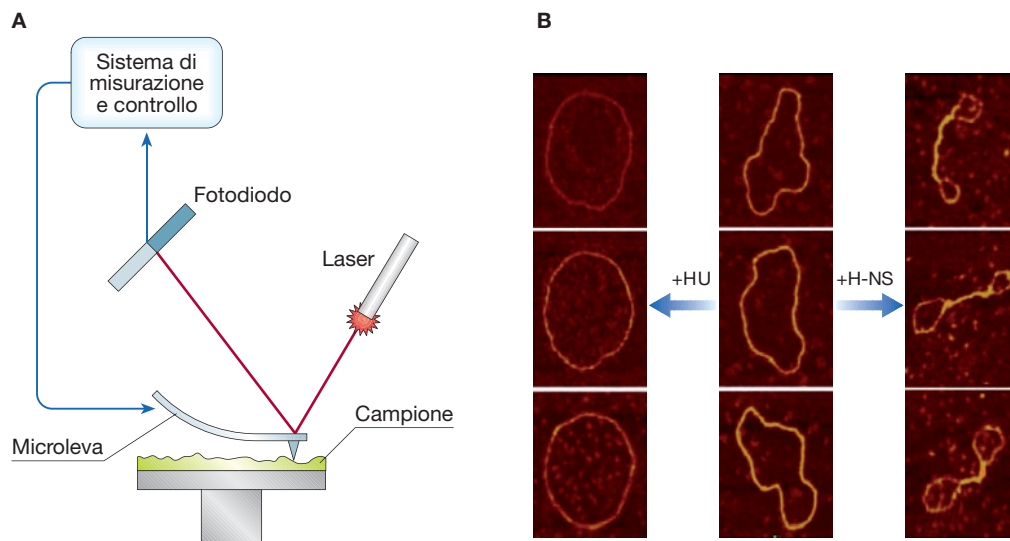


Figura 20.60 Microscopio a forza atomica (AFM). (A) Funzionamento schematico del microscopio a forza atomica. (B) Immagini ottenute con l'AFM che mostrano come il plasmide pUC19 nella sua forma topologicamente rilassata (pannelli centrali) venga aperto dall'interazione con la proteina batterica HU (1 dimer di

HU ogni 9 pb di pUC19; pannelli a sinistra), mentre viene compattato dall'interazione con la proteina batterica H-NS (1 dimer di H-NS ogni 12 pb di pUC19; pannelli a destra). [Riprodotta con il permesso della Federation of the European Biochemical Societies: Dame R.T., Goosen N. (2002) *FEBS Lett*, 529:151-156.]

mità della microleva verso una matrice di fotodiodi. Se la punta fosse utilizzata ad altezza costante, si correrebbe il rischio di toccare la superficie del complesso, danneggiandolo. Di conseguenza, nella maggior parte dei casi viene utilizzato un meccanismo di retroazione per regolare la distanza tra la punta e il campione al fine di mantenere costante la forza che agisce tra loro. Il risultato che si ottiene rappresenta la topografia della superficie del campione.

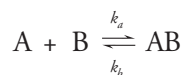
Interazioni proteina-proteina

Lo studio delle interazioni proteina-proteina è diventato di importanza fondamentale per comprendere la struttura e funzione delle numerose “macchine” proteiche coinvolte in tutti i processi molecolari della cellula.

Prima di descrivere alcuni degli approcci sperimentali e delle metodologie per lo studio delle interazioni proteina-proteina è importante avere almeno alcune sommarie informazioni sulle caratteristiche chimico-fisiche d'interazioni proteina-proteina, basate sull'analisi di un centinaio di complessi proteici analizzati e descritti nelle banche dati di proteine.

L'area di contatto tra due proteine è quasi sempre più grande di $1100 \text{ \AA}^2$. La formazione di un dimero coinvolge, mediamente, circa il 12% dell'area superficiale accessibile dei singoli monomeri, mentre tale percentuale sale al 17,4% e al 20,9%, per la formazione di un trimero o di un tetramero. Le forze che mediano le interazioni proteina-proteina includono interazioni elettrostatiche, legami idrogeno, forze di van der Waals e legami idrofobici. L'interfaccia di un complesso proteina-proteina non è, mediamente, meno polare o più idrofobico rispetto alla restante parte delle proteine che sono in contatto con il solvente: l'acqua è però normalmente esclusa dalla regione di contatto. È stato proposto che le forze idrofobiche guidino le interazioni proteina-proteina, mentre i legami idrogeno e i ponti salini contribuiscano di più alla specificità dell'interazione.

Le interazioni proteina-proteina possono essere descritte come delle semplici reazioni chimiche:



dove A e B sono le due proteine che interagiscono e AB è il complesso proteico. Complessi costituiti da molte proteine sono spesso considerati come se venissero assemblati attraverso l'aggiunta successiva delle varie subunità. Dal punto di vista qualitativo le interazioni proteina-proteina possono essere considerate “deboli” (di solito a vita breve) o “forti” (con una vita molto lunga).

Quantitativamente l'interazione tra due proteine segue la legge d'azione di massa

$$\frac{[A][B]}{[AB]} = \frac{1}{k_a} = K_d = \frac{k_d}{k_a}$$

dove:

- k_a = costante di velocità di secondo ordine per la reazione di associazione bi-molecolare;
- k_d = costante di velocità di primo ordine per la reazione di disassociazione uni-molecolare;
- $K_d = k_d/k_a$ = costante d'equilibrio per la disassociazione (K_a per l'associazione). K_d è correlata alle concentrazioni di A, B e AB all'equilibrio termodinamico. K_d ha le dimensioni di una concentrazione ed è espressa in moli/litro (o “M”). I valori di K_d osservati in reazioni biologicamente rilevanti sono molto ampi e si estendono su 12 ordini di grandezza, da 10^{-4} a 10^{-16} M.

Le interazioni proteina-proteina possono venire analizzate con metodi genetici, biochimici, molecolari e cellulari. Qui di seguito descriveremo brevemente alcuni dei metodi più utilizzati, alcuni dei quali (SPR e AFM) sono già stati descritti in quanto applicabili anche allo studio d'interazioni proteina-DNA.

■ Metodi genetici per lo studio d'interazioni proteina-proteina

In organismi a cui sono applicabili sofisticati approcci sperimentali di tipo genetico è possibile applicare strategie genetiche che individuano possibili interazioni proteina-proteina. Queste strategie hanno permesso di suggerire interazioni proteina-proteina non facilmente prevedibili e individuare interconnessioni insospettite tra pathway biochimici.

Tali metodologie si basano su approcci genetici classici e alcune di esse sono esemplificate in **figura 20.61**.

Per esempio, l'identificazione di **soppressori extragenici** della mutazione in uno specifico gene spesso rivela mutazioni in altri geni codificanti per proteine che interagiscono con la proteina originariamente mutata. In pratica, se una proteina X mutata non è più in grado d'interagire con un suo partner Y wild-type, è possibile che una mutazione nel gene codificante per la proteina Y sia in grado di compensare il difetto causato dalla mutazione in X (**fig. 20.61A**).

Un altro approccio genetico si basa sul fatto che la combinazione di due mutazioni può provocare un fenotipo molto più grave di quello osservato per le singole mutazioni: questo effetto viene chiamato “sintetico” e può essere così forte che la combinazione delle due mutazioni porta alla morte della cellula (si parla di **effetto letale sintetico**) (**fig. 20.61B**). Questo fenotipo è molto spesso causato da mutazioni in due geni che codificano per proteine che interagiscono tra di loro.

L'effetto di una mutazione nel gene codificante per una certa proteina può essere anche soppresso dalla sovrapproduzione di un'altra proteina che interagisce con la proteina mutata. Ciò si verifica se l'effetto della mutazione originaria diminuisce l'affinità di legame tra le due proteine: il fenotipo può, infatti, essere alleviato o eliminato se si aumenta la concentrazione della proteina interagente.

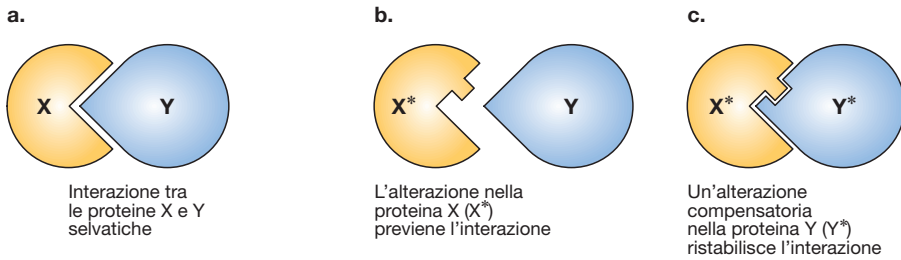
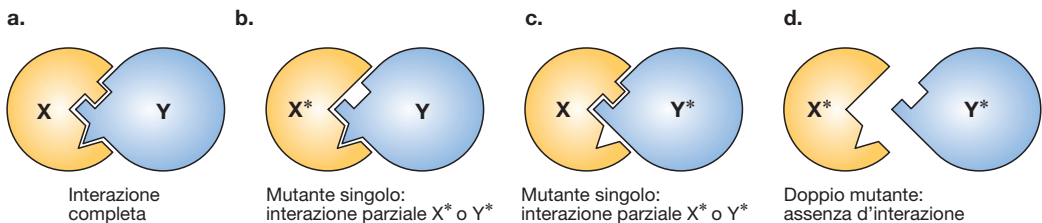
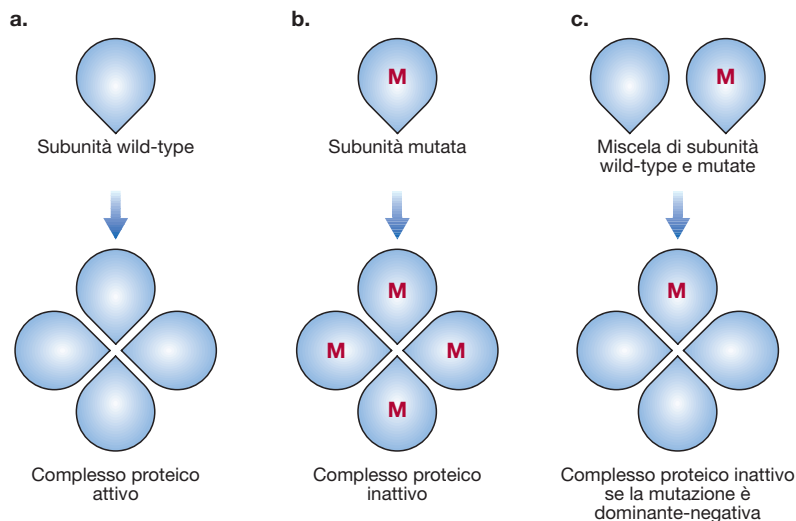
A. Soppressori extragenici**B. Effetto sintetico tra mutazioni****C. Effetto di mutazioni "dominanti negative"**

Figura 20.61 Approcci genetici per l'individuazione di geni co-dificanti per proteine che interagiscono tra di loro. L'utilizzo, i vantaggi e gli svantaggi degli approcci genetici per l'identificazione d'interazioni proteina-proteina sono discussi nel testo.

Va però considerato che la sovrapproduzione non è fisiologica e può alterare il bilancio d'interazioni tra i componenti di una macchina proteica. Può anche accadere (**fig. 20.61C**) che la proteina sovrapprodotta non sia funzionale e, in questo caso, può agire come una **mutazione** definita **dominante-negativa**.

Va ricordato che i metodi genetici sono estremamente utili soprattutto perché applicabili ad ampi screening indirizzati alla ricerca d'interazioni proteina-proteina ignote e non prevedibili o per verificare funzionalmente l'effetto fisiologico d'interazioni individuate con altri metodi. Va però anche detto che un'interazione identificata con approcci genetici dovrebbe sempre essere verificata anche con metodi biochimici e molecolari più diretti.

■ Metodi biochimici per lo studio d'interazioni proteina-proteina

Numerosi metodi biochimici classici sono utilizzati per studiare interazioni proteina-proteina e per identificare e purificare complessi proteici.

- **Co-purificazione.** Se due o più proteine interagiscono fortemente tra di loro è probabile che la loro interazione si mantenga durante vari passaggi di purificazione e, in tal caso, le proteine co-purificano insieme. Questo ha permesso la purificazione di molti complessi proteici stabili, alcuni dei quali sono stati descritti in altre parti del testo, quali le proteine multimeriche DNA polimerasi e RNA polimerasi.
- **Co-sedimentazione.** Se due o più proteine interagiscono tra di loro formano degli oggetti la cui massa è maggiore di quella dei singoli componenti. Ne deriva che un complesso proteico mostra una massa maggiore con tutte le metodologie biochimiche indirizzate a misurare le dimensioni di una macromolecola (centrifugazione, filtrazione su gel, elettroforesi in condizioni native ecc.).
- **Co-localizzazione.** Se due proteine interagiscono tra di loro ne consegue che la loro localizzazione all'interno della cellula sarà pressoché identica. Abbiamo già accennato che esistono tecnologie per attaccare a una proteina un epitopo riconosciuto da uno specifico anticorpo o da dei fluorocromi (proteine "etichettate" o *tagged*). La localizzazione di queste proteine modificate è, per esempio, più facilmente identificabile con tecniche di microscopia a fluorescenza, permettendo così di verificare se due o più proteine co-localizzano nello stesso compartimento della cellula.
- **Co-immunoprecipitazione.** Una proteina viene legata con altissima affinità da anticorpi che riconoscono uno o più epitopi presenti nella proteina in esame. Come abbiamo già descritto presentando la tecnica della ChIP, questi complessi proteina-anticorpo possono essere catturati o immunoprecipitati con varie tecniche. Se una proteina X interagisce con la proteina Y contro la quale sono disponibili degli anticorpi, sarà anch'essa presente nel complesso immunoprecipitato. La proteina legata all'anticorpo e quelle che co-immunoprecipitano con essa possono successivamente essere separate e visualizzate mediante elettroforesi in condizioni denaturanti.

Abbiamo già descritto altre potenti tecniche di purificazione di proteine (quale il *TAP-tagging*) che permettono di isolare proteine che interagiscono con la proteina-tagged. Anche queste proteine interagenti possono essere separate e visualizzate mediante elettroforesi. La parte complessa dell'intero approccio sperimentale consiste, però, nell'identificazione precisa delle proteine interagenti. Ciò si riesce a ottenere con una tecnica nota come spettrometria di massa.

La **spettrometria di massa** ha acquisito sempre più rilevanza con l'aumentare delle conoscenze sui genomi. Poiché oggi è nota la sequenza del genoma degli organismi preferenzialmente utilizzati nelle ricerche di Biologia Molecolare, nelle banche dati sono, di fatto, disponibili le sequenze amminoacidiche dedotte di tutti i geni potenzialmente codificanti per proteine. Il compito di identificare una o più proteine sconosciute si riduce, così, nel trovare una corrispondenza tra una sequenza anche parziale di una proteina ignota identificata sperimentalmente (ad es. mediante elettroforesi) e una sequenza amminoacidica presente nella banca dati genomica. Questo si ottiene combinando dati ottenuti tramite la spettrometria di massa con analisi bioinformatiche.

Particelle cariche hanno proprietà dinamiche ben definite quando soggette a campi elettrici e magnetici nel vuoto, e la spettrometria di massa sfrutta questa proprietà per separare ioni secondo il loro rapporto massa/carica. Tramite apparecchiature piuttosto complesse e costose si utilizza preferenzialmente una tecnica chiamata **MALDI-TOF** (*Matrix-Assisted Laser Desorption Ionization-Time Of Flight spectrometry*) per cui le proteine di un campione (ad es. una proteina separata mediante elettroforesi ed estratta dal gel) sono prima spezzate in frammenti che, dopo specifici trattamenti, sono depositi su una superficie di metallo. Il campione è poi colpito con un laser che provoca il distacco dei peptidi dalla

superficie sotto forma di un gas ionizzato in cui ciascuna molecola porta una o più cariche positive. I peptidi ionizzati sono accelerati in un campo elettrico e “volano” verso un rivelatore. Il tempo di volo fino al rivelatore dipende dalla loro massa e dalla loro carica: i peptidi più grossi si muovono più piano, quelli più carichi si muovono più veloci. Analizzando quei peptidi ionizzati che portano una singola carica, si possono determinare con estrema precisione le masse dei peptidi presenti nel campione originale. La tecnica è così precisa che, in opportune condizioni di analisi, si possono misurare variazioni di massa dell'ordine di 1 dalton o frazioni di esso (**fig. 20.62**).

Questa informazione viene ora usata per eseguire ricerche nelle banche dati per identificare quale sia la possibile proteina i cui frammenti sono compatibili con le masse dei peptidi individuati sperimentalmente mediante spettrometria di massa.

Impiegando due spettrometri di massa in tandem (una tecnica chiamata **MS/MS**) è anche possibile determinare la sequenza amminoacidica dei singoli peptidi nella miscela di peptidi della proteina in esame. Infatti, nel secondo spettrometro di massa un singolo peptide è spezzato in peptidi via via più corti che differiscono di un singolo amminoacido. Lo spettrometro separa questi peptidi

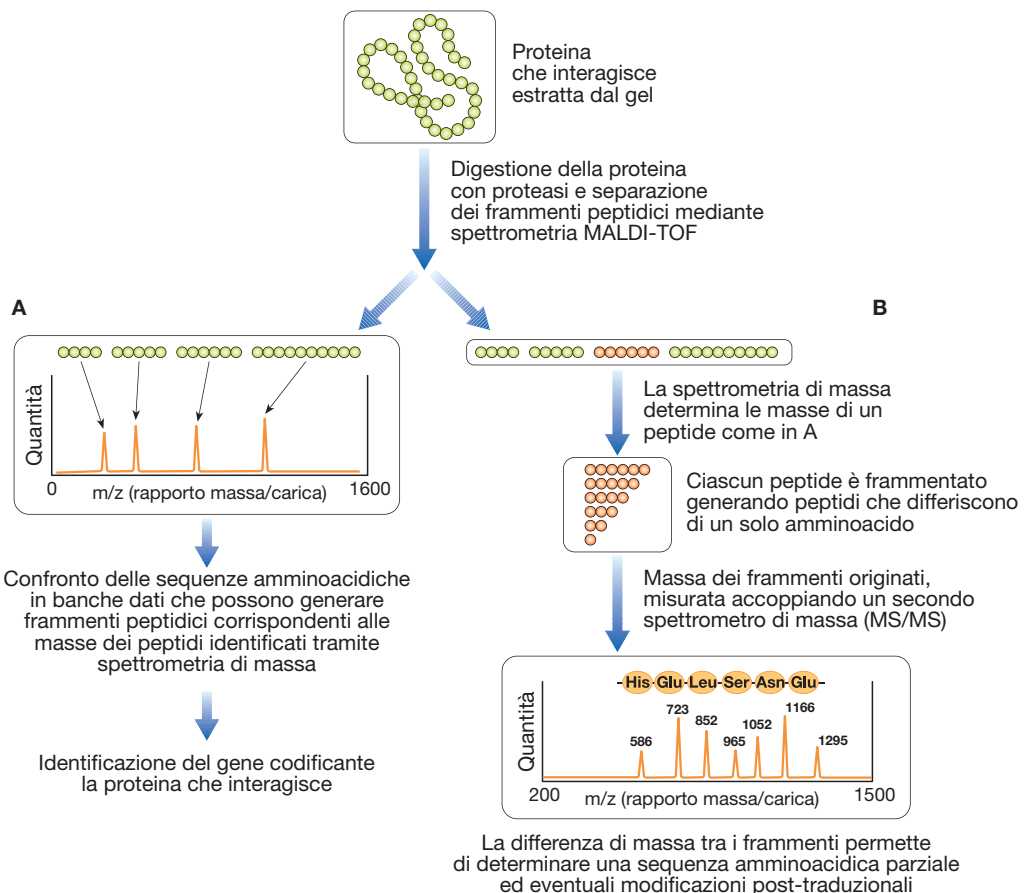


Figura 20.62 La spettrometria di massa permette d'identificare proteine che interagiscono tra di loro. (A) Una proteina interagente con un'altra proteina può essere separata e visualizzata mediante co-purificazione o co-immunoprecipitazione seguita da elettroforesi SDS-PAGE (vedi testo). La proteina interagente può essere identificata estraendo dal gel la proteina visualizzata tramite elettroforesi, sottoponendola poi a spettrometria di massa e integrando tale analisi con le banche dati genomiche. La proteina estratta dal gel è digerita con proteasi così da generare frammen-

ti peptidici che possono essere analizzati mediante la spettrometria MALDI-TOF (vedi testo), che, combinata con le banche dati di sequenze genomiche di proteine, permette d'identificare il gene codificante per la proteina interagente. (B) La spettrometria di massa, come descritto nel testo, può anche essere utilizzata (combinando due spettrometri di massa nella disposizione MS/MS) per determinare direttamente la sequenza amminoacidica dei frammenti peptidici, portando così all'identificazione di eventuali modificazioni post-traduzionali di specifici amminoacidi.

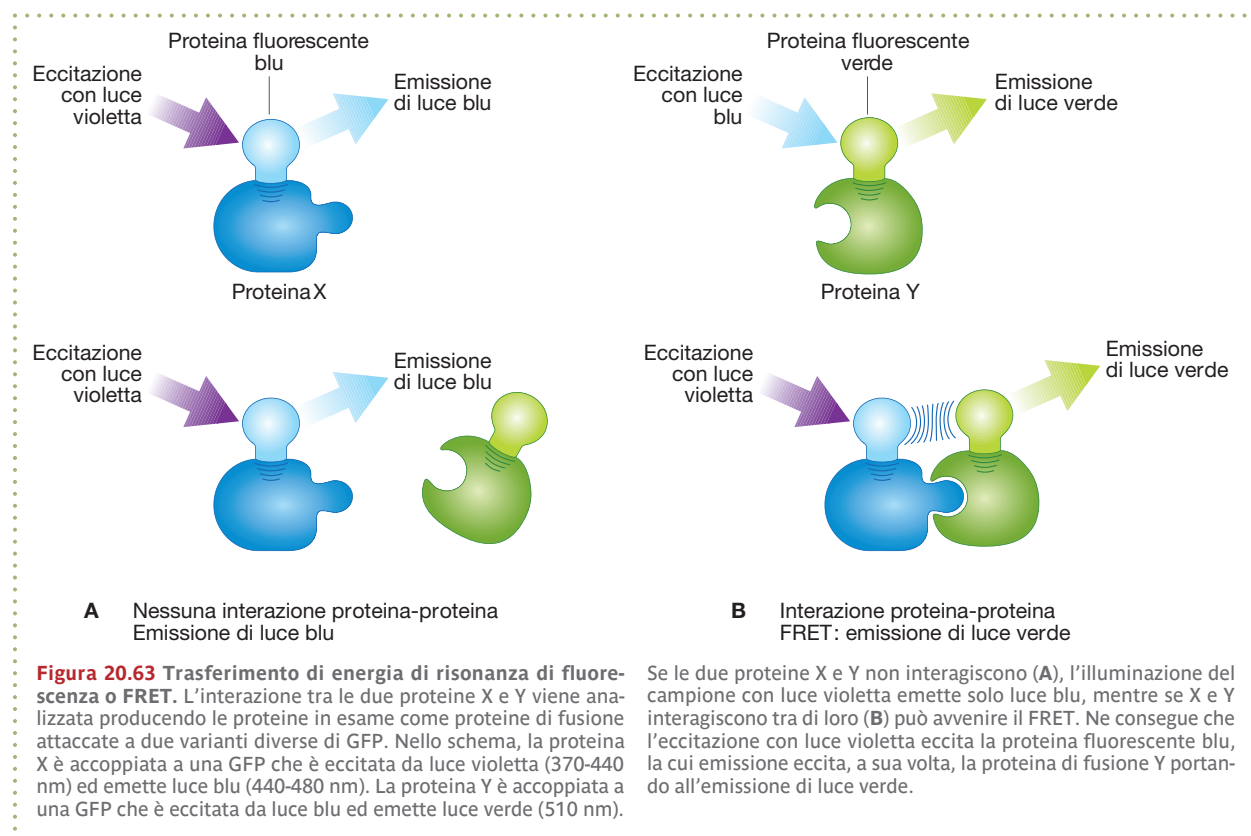
determinandone la massa; dalle differenze tra le loro masse si può dedurre la sequenza amminoacidica del peptide stesso (**fig. 20.62**). L'applicazione di questo secondo metodo di spettrometria è così precisa che è anche possibile determinare se i vari amminoacidi sono modificati post-traduzionalmente tramite fosforilazioni, acetilazioni, metilazioni ecc.

■ Combinazioni di analisi ottiche e metodologie di Biologia Molecolare

Negli ultimi anni sono state sviluppate metodologie innovative per identificare e studiare interazioni proteina-proteina che combinano metodi ottici e tecniche di biologia molecolare. Ne citeremo solo alcune e altre sono già state descritte poco sopra. Infatti, come accennato, la **risonanza plasmatica superficiale (SPR)** può essere utilizzata non solo per analizzare interazioni proteina-DNA ma anche interazioni tra proteine.

Un altro metodo ottico per valutare le interazioni di proteine fa uso delle proprietà di una **proteina fluorescente verde (GFP, Green Fluorescent Protein)**. La GFP è una proteina di 27 kDa espressa nella medusa *Aequorea victoria* che, se colpita da una radiazione di una certa lunghezza d'onda, è in grado di ri-emettere luce di colore verde acceso. La GFP è molto utilizzata per "etichettare" specifiche proteine mediante tecnologie classiche di DNA ricombinante, e tale "etichettatura" è utile per analizzare la localizzazione di una proteina nella cellula. Il gene codificante la proteina GFP selvatica può anche essere mutagenizzato, così da produrre forme modificate di GFP in grado di assorbire radiazioni diverse da quella originaria e di ri-emettere colorazioni diverse dal verde.

In una metodologia che permette di analizzare l'interazione tra due proteine (chiamata **trasferimento di energia di risonanza di fluorescenza o FRET**) le due proteine in esame sono marcate ciascuna con un fluorocromo in grado di emettere luce di diverso colore (**fig. 20.63**). Se le due proteine con attaccati i due



Se le due proteine X e Y non interagiscono (A), l'illuminazione del campione con luce violetta emette solo luce blu, mentre se X e Y interagiscono tra di loro (B) può avvenire il FRET. Ne consegue che l'eccitazione con luce violetta eccita la proteina fluorescente blu, la cui emissione eccita, a sua volta, la proteina di fusione Y portando all'emissione di luce verde.

fluorocromi diversi si trovano molto vicine fra loro (entro 1-10 nm), l'energia della luce assorbita verrà trasferita da un fluorocromo all'altro, così che l'interazione tra le due proteine può venire visualizzata eccitando il primo fluorocromo con una specifica lunghezza d'onda e misurando il colore d'emmissione del secondo fluorocromo. Tale tecnica, oltre a stabilire l'interazione tra le due proteine, può permettere d'individuare dove l'interazione avviene all'interno della cellula.

■ Tecnica del doppio ibrido

La tecnica denominata del **doppio ibrido** o **Two Hybrid (TH)** è una metodologia per individuare e analizzare interazioni proteina-proteina che ha trovato ampia applicazione dopo la scoperta che i fattori trascrizionali sono proteine modulari contenenti un dominio di legame al DNA (*DNA Binding Domain* o DBD) e un dominio di attivazione trascrizionale (*Activating Domain* o AD).

Infatti, nel 1985 Roger Brent e Mark Ptashne dimostrarono che proteine di fusione composte dal DBD di un fattore trascrizionale e dall'AD di un altro fattore trascrizionale sono in grado di attivare la trascrizione. Questo esperimento dimostrò che per indurre la trascrizione di un gene i due domini DBD e AD possono funzionare indipendentemente l'uno dall'altro. Questo concetto fu applicato da Stanley Fields e collaboratori, che nel 1989 svilupparono la tecnica del doppio ibrido e ne individuarono le applicazioni. Uno schema generale del doppio ibrido è mostrato in **figura 20.64A**.

Il dominio di legame al DNA (DBD) di un fattore trascrizionale è fuso alla proteina X. Tale proteina ibrida è in grado di legarsi in corrispondenza della sequenza di DNA riconosciuta dal DBD che si trova in prossimità di un "gene

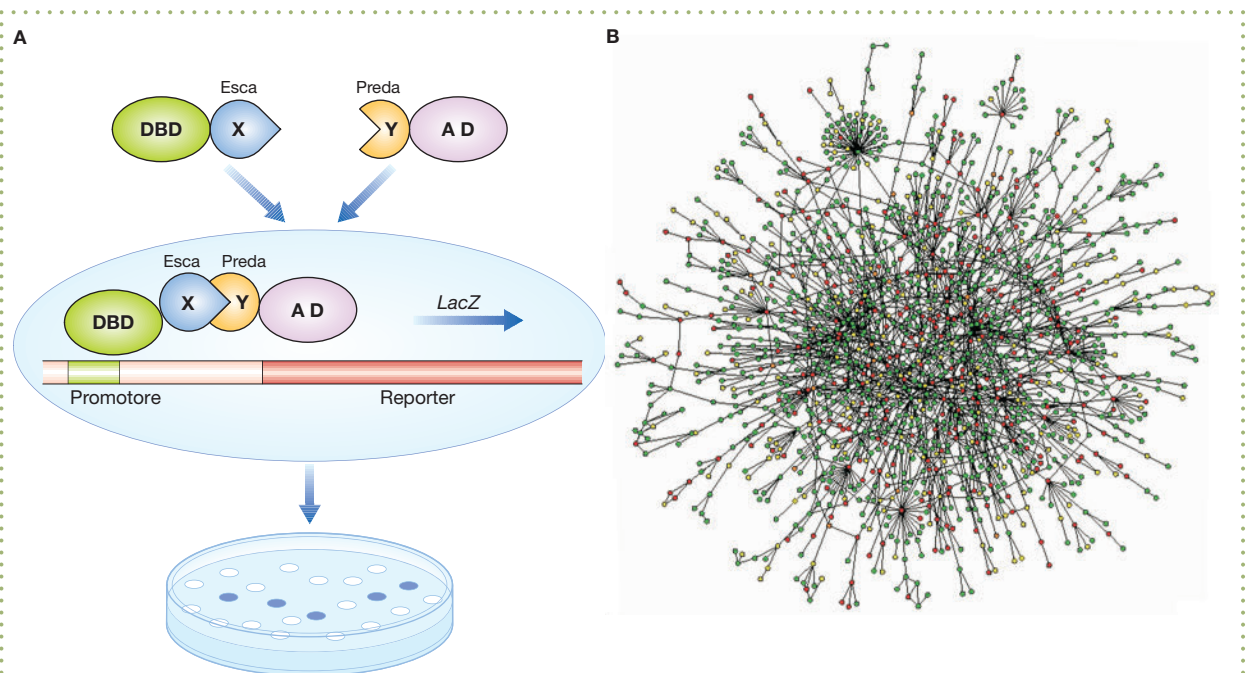


Figura 20.64 Tecnica del doppio ibrido (TH). (A) Come spiegato più estesamente nel testo, viene eseguita una doppia trasformazione di cellule di lievito con due plasmidi: il primo esprime il gene codificante la proteina di fusione X-DBD, proteina X fusa al dominio di legame al DNA che riconosce e lega il promotore che si trova a monte del gene reporter; il secondo plasmide esprime la proteina Y-AD, proteina Y fusa al dominio di attivazione della trascrizione. Solo se le due proteine X e Y interagiscono tra loro, le cellule di lievito trascriveranno il gene reporter *LacZ*. La facilità

di analisi e selezione di questo reporter permette l'applicazione di tale metodologia su grande scala ad ampi screening genetici tesi a individuare tutti i possibili interattori di una proteina. (B) Screening sistematici tramite doppio ibrido applicati a molte, talvolta a tutte le proteine di una cellula hanno portato a identificare la complessa rete d'interazioni proteina-proteina nella cellula stessa. Le palline colorate rappresentano le diverse proteine e ciascuna delle linee che nello schema congiungono le diverse proteine indica un'interazione proteina-proteina stabilita mediante TH.

reporter” ma, mancando di un dominio di attivazione trascrizionale (AD), non è in grado di attivarne la trascrizione. Analogamente, una proteina di fusione costituita dalla proteina Y legata all’AD dello stesso o di un altro fattore trascrizionale, non sarà in grado di attivare la trascrizione del gene reporter in quanto incapace di legarsi al DNA. Tuttavia, se la proteina di fusione X interagisce con la proteina di fusione Y, l’interazione fisica tra le due proteine è in grado di ricostituire un fattore trascrizionale attivo in cui l’azione combinata del DBD e dell’AD è resa possibile dall’interazione tra X e Y.

Esistono numerosi varianti del TH che fanno uso di DBD e AD di vari fattori trascrizionali di procarioti ed eucarioti e che hanno svariate applicazioni. Qui vogliamo solo ricordare che il doppio ibrido non solo indica che due proteine specifiche, quali X e Y, interagiscono tra di loro, ma può essere applicato alla ricerca di interattori sconosciuti di una specifica proteina. In tali screening genetici la proteina di fusione X contenente il DBD è considerata un’*esca* che viene utilizzata per lo screening di una libreria di cDNA in cui vengono preparate fusioni tra le sequenze codificanti dei cDNA e il dominio AD; i cDNA vengono considerati come delle *prede*. L’interazione tra l’esca e una preda (cioè un cDNA che codifica per una proteina che interagisce con la proteina X) induce la trascrizione del gene reporter; l’espressione di tale gene, per esempio il gene *lacZ* di *E. coli*, può essere facilmente individuata con la tecnica del “bianco-blu” (**fig. 20.18**) e, quindi, è applicabile a screening su larga scala (**fig. 20.64B**).

Screening sistematici tramite doppio ibrido applicati a molte e talvolta a tutte le proteine di una cellula hanno portato a identificare la complessa rete d’interazioni proteina-proteina nella cellula stessa (**fig. 20.64C**). Va sottolineato che il TH è normalmente eseguito all’interno di cellule di lievito, anche se le proteine in esame possono essere codificate da genomi diversi dal lievito stesso; il TH, quindi, individua interazioni proteina-proteina *in vivo*. Le interazioni suggerite dal TH devono essere verificate con altri metodi biochimici più diretti; infatti, anche se la maggior parte delle interazioni indicate dal TH si sono dimostrate veritiere, in alcuni casi la tecnica ha evidenziato interazioni che non è stato possibile confermare con altri metodi.

Lecture di approfondimento

- Auerbach, D., Thamiy, S., Hottiher, M.O. e Staglijar, I. (2002), «The post-genomic era of interactive proteomics: facts and perspectives», *Proteomics*, **2**, pp. 611-623.
- Bilibewski, U. (2009), «DNA microarrays: an introduction to the technology», *Methods in Molecular Biology*, **36**, pp. 758-765.
- Kapanidis, A.N. e Strick, T. (2009), «Biology, one molecule at a time», *Trends in Biochemical Sciences*, **34**, pp. 234-243.
- Mardis, E.R. (2008), «Next generation DNA sequencing methods», *Annual Review of Genomics and Human Genetics*, **9**, pp. 387-402.
- Sambrook, J. e Russell, D.W. (2001), *Molecular cloning: a laboratory manual*, III ed., Cold Spring Harbor Press.

21

Bioinformatica e Genomica

21.1 Sequenziamento e assemblaggio di genomi completi

21.2 Sequenziamento del trascrittoma

21.3 Identificazione e annotazione di geni

21.4 Confronto e allineamento di sequenze

21.5 Identificazione e annotazione di regioni regolatorie e caratterizzazione delle proprietà epigenetiche della cromatina

21.6 Annotazione delle sequenze proteiche

21.7 Banche dati e browser genomici

21.8 Evoluzione molecolare

I grandi progressi della Biologia Molecolare sono stati resi possibili dal contemporaneo sviluppo di nuove straordinarie tecniche, già descritte nel capitolo 20, che hanno ampliato in modo spettacolare le possibilità di indagine sperimentale. Queste tecniche sono oggi in grado di produrre una vastissima mole di dati e informazioni, di natura eterogenea, che vanno dalle sequenze degli acidi nucleici e delle proteine ai profili di espressione genica, alle interazioni proteine-acidi nucleici. Questi dati non avrebbero alcun valore senza strumenti appropriati per la loro gestione e analisi. A questo scopo, alla fine degli anni settanta, in concomitanza con lo sviluppo delle prime tecniche di sequenziamento degli acidi nucleici, è nata la Bioinformatica, una disciplina che in questi ultimi anni sta vivendo uno sviluppo eccezionale. Infatti, si stima che, in prospettiva, ogni laboratorio avanzato di ricerca in ambito biomolecolare dovrebbe disporre di un numero di bioinformatici almeno confrontabile a quello dei ricercatori tradizionali.

La Bioinformatica necessita di piattaforme hardware dotate di elevata capacità di storage e di calcolo, che difficilmente possono essere nella disponibilità di singoli laboratori di ricerca. Per questa ragione sono sorti grandi centri a livello internazionale, come l'European Bioinformatics Institute (EBI) di Cambridge (UK) e il National Centre of Biological Information (NCBI) di Bethesda (USA), che forniscono accesso a numerose banche dati e servizi di analisi bioinformatica.

Nei paragrafi che seguono illustreremo alcune delle applicazioni bioinformatiche più rilevanti.

21.1 Sequenziamento e assemblaggio di genomi completi

I recenti progressi nelle strategie e nelle tecnologie per il sequenziamento del DNA stanno operando una profonda rivoluzione in moltissimi campi, parti-

colarmente nel settore biomedico, di cui ancora non si riesce a valutare l'intera portata e le implicazioni pratiche nella nostra vita quotidiana.

A partire dal 1995, quando è stato pubblicato il primo genoma completo di un organismo vivente (il batterio *Haemophilus influenzae*), vi sono stati notevolissimi avanzamenti sul piano sperimentale e tecnologico che hanno permesso il sequenziamento completo di un gran numero di genomi procariotici ed eucariotici.

In particolare, il **progetto "Genoma Umano"**, avviato nel 1990, ha avuto un ruolo fondamentale nel favorire tutta una serie di importanti innovazioni che hanno reso sempre più veloce ed economico il sequenziamento genomico. Sono stati introdotti decisivi miglioramenti nella tecnica di sequenziamento basata sul metodo di Sanger (*vedi* par. 20.7), come lo sviluppo di sistemi di marcatura fluorescente e della tecnologia laser per il rilevamento delle molecole di DNA, della strumentazione per l'elettroforesi capillare, l'introduzione di nuove DNA polimerasi termostabili e processive. In questo modo è stato possibile ottenere sequenze di alta qualità di lunghezza fino a circa 1 kpb.

Inoltre, sono stati sviluppati sistemi robotizzati per automatizzare molte delle operazioni richieste nelle diverse fasi del sequenziamento, come la costruzione delle librerie batteriche, la purificazione del DNA estratto da ciascun clone, le reazioni di sequenziamento e il caricamento dei campioni negli apparecchi di sequenziamento automatico.

Oggi, con l'avvento di strumenti di seconda e terza generazione (*vedi* par. 20.8) la capacità di sequenziamento è ulteriormente aumentata di alcuni ordini di grandezza (sino ad alcune migliaia di Gpb per corsa) e, contestualmente, si è realizzata una drastica riduzione dei costi. Si stima che il progetto "Genoma Umano" sia costato circa 3 miliardi di dollari. È stata recentemente raggiunta la soglia di 1000 dollari per il costo del sequenziamento dell'intero genoma di un individuo e si stima che a breve verrà raggiunta quella dei 100 dollari rendendo la determinazione del profilo genetico di ogni individuo una banale routine, anche se con implicazioni sia cliniche che etiche di grandissima rilevanza.

I problemi legati al sequenziamento di un genoma più o meno complesso non sono relativi alla produzione delle sequenze in sé, ma alla determinazione dell'ordine con cui queste sequenze devono essere disposte per ottenere la conoscenza completa del genoma, ovvero del corretto "assemblaggio" dell'intero genoma. Il problema deriva dal fatto che ogni singola sequenza prodotta con le attuali metodologie è molto più corta della dimensione del genoma da cui proviene. Il problema permane (anche se molto attenuato) con i sequenziatori di terza generazione descritti nel paragrafo 20.8 che, come abbiamo visto, possono produrre read molto lunghe (>20 kpb). Se consideriamo che anche il più piccolo genoma batterico ha la dimensione di alcune centinaia di migliaia di pb, è evidente che le singole sequenze (la cui lunghezza è dell'ordine delle centinaia di pb per i sistemi di seconda generazione e delle migliaia di pb per quelli di terza) rappresentano solo piccoli frammenti del genoma. La ricostruzione della sequenza genomica consiste nella disposizione ordinata di tali frammenti sino a coprire, sperabilmente, l'intero genoma. Il problema è simile alla soluzione di un grande puzzle in cui tutti i frammenti devono essere collocati nella giusta posizione per ricostruire l'immagine riportata sulla scatola. La disposizione ordinata dei frammenti di sequenza per la ricostruzione del genoma è, però, molto più complessa della ricostruzione di un puzzle, a causa della presenza nel genoma di sequenze ripetute; pertanto, contrariamente a quanto accade per un puzzle, frammenti di sequenza virtualmente identici possono provenire da regioni differenti del genoma. Inoltre, non è a disposizione l'immagine integrale riportata sulla scatola per guidare il processo di ricostruzione. Ne consegue che, se vogliamo ricostruire una sequenza genomica completa, abbiamo bisogno di un insieme di "marcatori", ovvero di elementi genetici unici di cui sia nota con precisione la disposizione lungo il genoma.

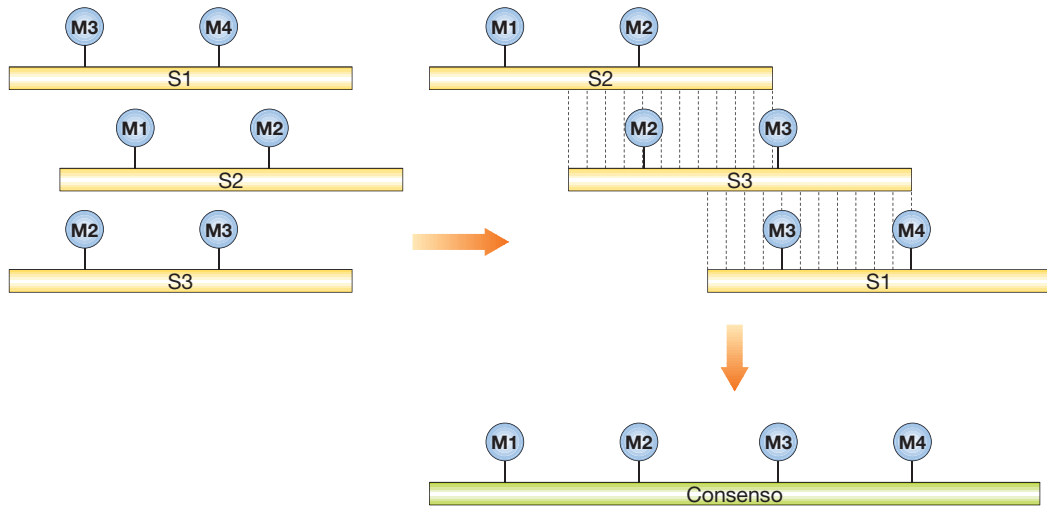


Figura 21.1 Assemblaggio di sequenze per ottenere una sequenza consenso. Date tre sequenze (S1, S2 e S3) e quattro marcatori (M1, M2, M3 e M4), possiamo disporre nell'ordine corretto le tre sequenze sulla base della localizzazione dei marcatori in comune.

L'insieme di questi marcatori costituisce quella che viene denominata **mappa fisica** o **mappa genetica**, in funzione delle modalità attraverso le quali viene determinata la disposizione spaziale relativa dei marcatori (per ulteriori dettagli consultare un buon testo di Genetica). Più elevato è il numero di marcatori di cui si dispone, maggiore è la risoluzione della mappa e, di conseguenza, tanto più agevole sarà la ricostruzione dell'intera sequenza. Supponiamo, per esempio, di avere una mappa costituita da una serie ordinata di marcatori M1, M2, M3 ecc. Se disponiamo di tre frammenti di sequenza S1, S2 e S3 che contengono rispettivamente i marcatori M3-M4, M1-M2 e M2-M3, potremo ordinare in modo non ambiguo le sequenze come S2-S3-S1 per costruire una sequenza più lunga, detta **sequenza consenso**, data dalla somma delle lunghezze dei singoli frammenti meno la lunghezza delle regioni di sovrapposizione.

Come si può vedere dall'esempio riportato in **figura 21.1**, la ricostruzione della sequenza "assemblata" richiede che sequenze distinte condividano uno o più marcatori, ovvero che ci siano regioni di sovrapposizione più o meno estese tra le sequenze prodotte. Questo vuol dire che se vogliamo sequenziare un genoma batterico di 5 Mpb, la quantità di sequenze che sarà necessario produrre deve essere molto maggiore. Il rapporto tra la quantità totale delle sequenze prodotte, che possiamo decidere in partenza, e la dimensione del genoma definisce il grado di copertura o "coverage" del nostro progetto di sequenziamento. Per esempio, se nel caso riportato sopra la quantità totale di sequenze prodotte è pari a 50 Mpb, avremo realizzato una copertura di 10 volte, detta anche 10×. Il grado di copertura necessario a completare la sequenza di un determinato genoma non è costante, ma dipende dalle sue dimensioni e dal grado di ripetitività oltre che dalla lunghezza e dalla natura (single- o paired-end, vedi par. 20.8) delle sequenze generate dall'apparato di sequenziamento. Il grado di copertura necessario a raggiungere un determinato livello di completezza sarà maggiore per genomi altamente ripetitivi e con unità ripetute di lunghezza variabile. Alcune porzioni del genoma, come le regioni centromeriche che possono estendersi per molte megabasi e che sono costituite da unità ripetute di media lunghezza, non possono essere sequenziate compiutamente con le attuali tecniche.

Il DNA genomico non può essere sequenziato direttamente ma prima deve essere frammentato, e i frammenti risultanti, in funzione dell'approccio adottato, eventualmente clonati in vettori appropriati. I vettori più utilizzati a que-

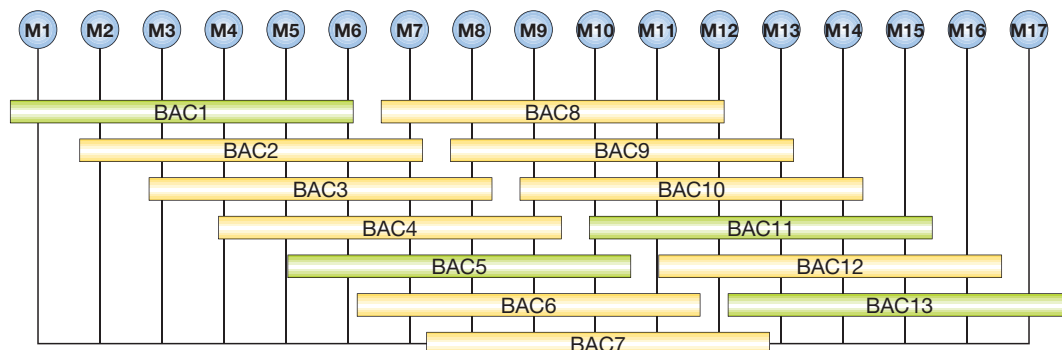


Figura 21.2 Minimum Tiling Path (MTP). È dato dal numero minimo di cloni in grado di coprire un intero genoma (cromosoma). Nell'esempio in figura, questo è costituito da 4 cloni su 13 (in verde).

sto scopo sono i **BAC** (*Bacterial Artificial Chromosomes*), che possono contenere frammenti di DNA piuttosto grandi (100-200 kpb) e possono essere propagati in *Escherichia coli*. Quindi, il materiale di partenza per un progetto di sequenziamento genomico è, normalmente, una libreria di cloni BAC contenenti una vasta collezione di inserti che statisticamente coprono più volte tutto il genoma.

A questo punto è possibile seguire due strategie distinte, denominate rispettivamente **Clone-by-clone shotgun Genome Sequencing (CGS)** e **Whole Genome shotgun Sequencing (WGS)**.

Nel primo caso è necessario il clonaggio e i cloni BAC devono essere in primo luogo disposti in modo ordinato lungo il genoma (o i diversi cromosomi) in modo da determinare il numero minimo di cloni in grado di garantire la copertura completa (**fig. 21.2**). A questo punto, i cloni prescelti, che costituiscono quello che viene denominato **Minimum Tiling Path (MTP)**, possono essere sequenziati con l'approccio *shotgun*. Ovvero, gli inserti presenti nei diversi cloni sono frammentati in segmenti di 2-10 kpb e subclonati in altri vettori plasmidici o fagici. Si procede, quindi, al sequenziamento degli inserti di questi subcloni, utilizzando i primer universali specifici del vettore di clonaggio usato, a entrambe le estremità. Il sequenziamento delle estremità dell'inserto è molto più semplice e automatizzabile, in quanto vengono utilizzati i primer universali specifici del vettore per tutte le reazioni. Le sequenze così ottenute, disposte in modo casuale lungo il clone BAC considerato, vengono poi assemblate sino a completare la sequenza dell'intero inserto. Il processo è ripetuto per tutti i cloni del MTP e le sequenze risultanti vengono assemblate per completare la sequenza del genoma (cromosoma) (**fig. 21.3A**).

La strategia alternativa, nota come **Whole Genome Shotgun (WGS)**, non richiede il clonaggio e salta completamente la fase della costruzione del MTP, in quanto realizza un sequenziamento casuale degli inserti prodotti (**fig. 21.3B**). Questa strategia è molto più semplice e veloce, ma rende la ricostruzione della sequenza genomica più complessa, soprattutto nel caso di genomi ad alto contenuto di DNA ripetitivo e non codificante. Il progetto Genoma Umano finanziato con fondi pubblici è stato condotto con la strategia CGS, richiedendo un tempo molto maggiore di quello che è stato necessario alla compagnia privata Celera Genomics, che invece ha adottato l'approccio WGS. Nel 2001 sono state pubblicate quasi simultaneamente, sulle riviste *Nature* e *Science*, due versioni del genoma umano, da parte del consorzio pubblico e da parte di Celera Genomics. Lo sviluppo di algoritmi e software di assemblaggio sempre più efficienti ha prodotto una progressiva affermazione dell'approccio WGS, che di fatto ha completamente soppiantato la strategia CGS. L'assemblaggio delle sequenze prodotte in un progetto WGS è possibile in quanto, per il clonaggio, possono essere usati vettori diversi che possono contenere inserti

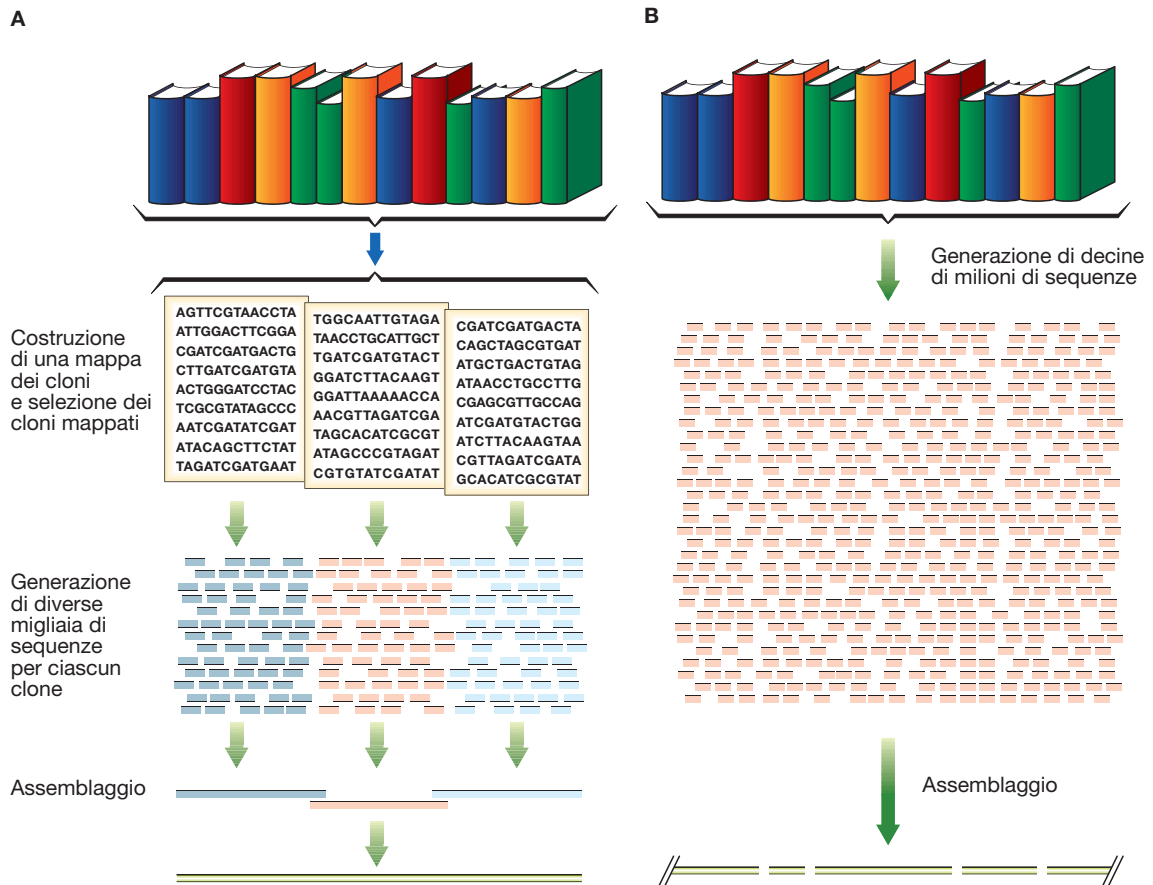


Figura 21.3 Strategie CGS e WGS per il sequenziamento dei genomi. (A) Rappresentazione schematica della procedura di sequenziamento *clone-by-clone*. Il genoma è rappresentato come se fosse un'enciclopedia i cui singoli volumi indicano i diversi cromosomi. La mappa fisica del genoma risulta fondamentale per ordinare la serie dei cloni e determinare quelli che costituiscono il Minimum Tiling Path. Ciascun clone, che possiamo immaginare corrispondere a una pagina del volume, viene sottoposto al sequenziamento

shotgun. Le sequenze risultanti vengono utilizzate per costruire la sequenza consenso di ciascun clone. Tutte le sequenze consenso ottenute vengono assemblate per costituire la sequenza consenso di ciascun cromosoma. (B) Rappresentazione schematica della procedura di sequenziamento Whole Genome Shotgun. In questo caso, non viene utilizzata la mappa fisica per costruire il Minimum Tiling Path e si procede direttamente all'assemblaggio delle sequenze dei cromosomi.

di dimensioni molto variabili, da 1-2 kpb per i vettori plasmidici a oltre 100 kpb per i vettori BAC, di cui vengono sequenziate entrambe le estremità. Queste estremità, se corrispondenti a sequenze uniche, attraverso l'applicazione di programmi bioinformatici piuttosto sofisticati, consentono la ricostruzione affidabile della sequenza genomica di interi cromosomi (**fig. 21.4**). Naturalmente, possono rimanere dei buchi, più o meno estesi, in funzione del grado di copertura ottenuto e della ripetitività del genoma.

L'avvento delle piattaforme di sequenziamento di nuova generazione (*vedi* par. 20.8) consente di eludere anche la laboriosa fase di clonaggio, rendendo quindi l'intero processo sensibilmente più veloce ed economico. Inoltre, la possibilità di ottenere letture di sequenza molto lunghe, fino a 20 kpb, come quelle prodotte dalle piattaforme Pacific Bioscience e Oxford Nanopore, rende molto più accurato, veloce ed economico il processo di assemblaggio del genoma.

21.2 Sequenziamento del trascrittoma

Il sequenziamento del genoma è il primo passo, necessario ma non sufficiente, per decodificare l'informazione genetica in esso depositata. Normalmente, l'in-

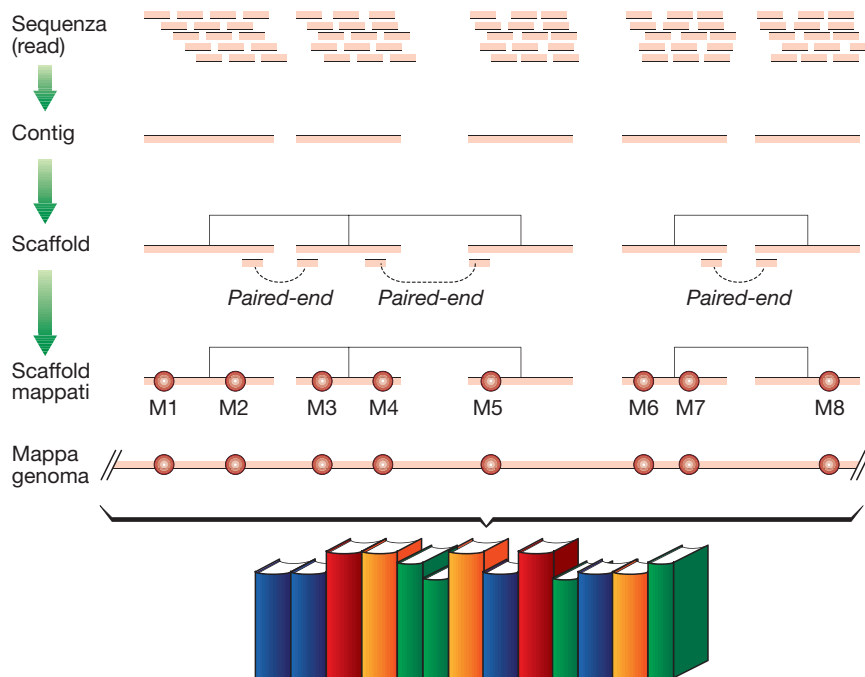


Figura 21.4 Strategia di assemblaggio basata sull'informazione dei "paired-end". Le sequenze singole generate dal sequenziamento WGS vengono assemblate in contig sulla base della sovrapposizione di sequenze uniche (per "contig" si intende una porzione continua di sequenza genomica ottenuta dall'assemblaggio di un insieme di letture di sequenze sovrapposte). I contig così ottenuti possono a loro volta essere ordinati per formare scaffold sulla base dell'informazione di connettività fornita dalle coppie di sequenze che derivano dalle estremità degli inserti di uno stesso clone (*paired-end*). Infatti, se le

due porzioni di una lettura di sequenza *paired-end* mappano su due contig diversi, possiamo desumere che essi siano contigui nel genoma. L'affidabilità degli scaffold così ottenuti può essere controllata mappando su di essi eventuali marcatori noti (pallini colorati) e verificandone la congruenza. A differenza del contig, lo scaffold è una porzione di sequenza genomica che presenta dei buchi o gap. La mappa completa del genoma può successivamente essere ottenuta attraverso un processo di finitura che comunque può lasciare "buchi" più o meno estesi.

formazione genetica viene espressa attraverso la trascrizione dei geni in RNA e il successivo processo di traduzione degli RNA messaggeri. Tuttavia, mentre l'informazione genetica racchiusa nel genoma è la stessa (con qualche eccezione) in tutte le cellule di uno stesso organismo, ognuna di queste può essere caratterizzata da un programma genetico specifico, in funzione della condizione fisiologica o patologica, dello stadio di sviluppo, o del tipo di cellula o compartimento subcellulare, tessuto od organo cui essa appartiene. Per esempio, una cellula epatica avrà un programma genetico diverso da quello di una cellula renale, che si traduce nel fatto che esse avranno un differente profilo di espressione, determinato dal sottoinsieme di geni espressi e dal loro livello quantitativo di espressione. Solo un numero piuttosto limitato di geni, detti costitutivi o housekeeping, viene espresso in modo ubiquitario in tutti i tipi di cellule. Ma come si fa a determinare il profilo di espressione, ovvero il programma genetico, dei diversi tipi cellulari e in diverse condizioni? È evidente che per affrontare questo problema è necessario condurre, in parallelo a ogni progetto di sequenziamento del genoma, un progetto di sequenziamento del **trascrittoma**, ovvero dell'insieme di molecole di RNA presenti nei diversi tipi cellulari in esame, nelle diverse possibili condizioni.

Si può facilmente comprendere come lo studio del trascrittoma sia molto più complesso di quello del genoma, in quanto a uno stesso genoma corrispondono, in pratica, innumerevoli trascrittomi, ciascuno specifico per una determinata condizione. Il sequenziamento dell'RNA richiede che questo venga prima retro-trascritto in cDNA (*vedi* par. 20.5).

Vi sono varie modalità sperimentali per la costruzione di librerie di cDNA. In particolare, nelle librerie rappresentative la frequenza di un determinato trascrit-

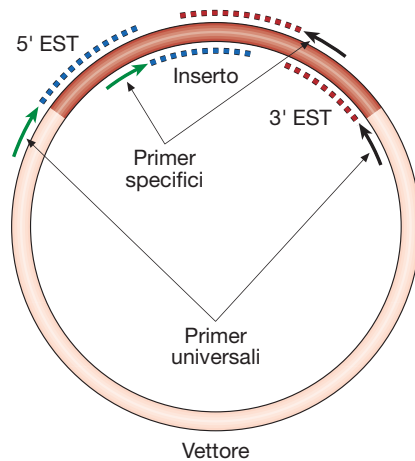


Figura 21.5 Strategie per la generazione di sequenze EST e trascritti "full-length". Le sequenze EST 5' e 3' possono essere generate utilizzando primer specifici del vettore. In questo modo non è normalmente possibile ottenere l'intera sequenza dell'inserto (*full-length*), che, invece, può essere completata utilizzando una o più coppie di primer inserto-specifiche.

to è proporzionale alla sua reale abbondanza nella cellula, mentre nelle librerie normalizzate si realizza un depauperamento dei trascritti più abbondanti e un arricchimento di quelli più rari, che magari non verrebbero rivelati nelle librerie rappresentative. I trascritti poco abbondanti generalmente codificano per proteine regolative, come i fattori di trascrizione, e svolgono una funzione molto importante nella fisiologia della cellula.

Un altro aspetto rilevante riguarda la natura dell'RNA di partenza utilizzato per la costruzione della libreria di cDNA. Infatti, è possibile usare l'RNA totale oppure esclusivamente la frazione poliadenilata. Nel secondo caso vengono privilegiati gli mRNA codificanti proteine, che generalmente costituiscono meno del 5% dell'RNA totale di una cellula, ma non si ottengono informazioni relative alla frazione non poliadenilata dell'RNA, generalmente non codificante proteine, che sta assumendo un ruolo sempre più importante nei processi di regolazione genica. D'altra parte, se si utilizza l'RNA totale, questo sarà costituito per la maggior parte da RNA ribosomale, che, da solo, rappresenta oltre il 80% dell'RNA totale, a meno che non vengano adottati protocolli sperimentali per la rimozione selettiva dell'rRNA.

La libreria di cDNA prodotta deve essere poi clonata nel caso di sequenziamento con il tradizionale metodo di Sanger oppure direttamente utilizzata per la preparazione delle librerie per il sequenziamento massivo con le piattaforme NGS.

La strategia di sequenziamento con il **metodo di Sanger** è analoga a quella precedentemente illustrata per il sequenziamento genomico, in quanto vengono utilizzati dei primer universali, specifici del vettore utilizzato, per determinare la sequenza nucleotidica a entrambe le estremità degli inserti presenti nella libreria di cloni di cDNA precedentemente costruita (**fig. 21.5**). Le sequenze così ottenute, della lunghezza massima di 1 kpb, corrispondono a frammenti di trascritti maturi e vengono quindi denominate **sequenze EST** (*Expressed Sequence Tags*). Le sequenze EST sinora prodotte sono depositate in un'apposita banca dati, denominata **dbEST** e accessibile in rete dai principali centri di bioinformatica internazionali. In questa banca dati sono contenute decine di milioni di sequenze, per molti organismi, derivate da diversi tessuti in condizioni normali e patologiche (**tab. 21.1**).

L'avvento delle piattaforme NGS ha di fatto superato le strategie tradizionali per il sequenziamento delle EST che comprendevano la fase di clonaggio. La strategia attualmente utilizzata, denominata **RNA-Seq**, prevede la trasformazione in cDNA di una preparazione di RNA, la generazione di frammenti di cDNA di lunghezza omogenea, cui vengono poi aggiunti degli adattatori specifici. La libreria di frammenti così generata viene poi direttamente sottoposta al sequenziamento massivo. È così possibile ottenere le sequenze di una o entrambe le estremità dei frammenti (*single- o paired-end sequencing*) da diversi tipi di preparazioni di

Tabella 21.1 Alcune delle sequenze EST depositate all'European Bioinformatics Institute (dati gennaio 2017).

Specie	Numero di EST
<i>Homo sapiens</i>	8 704 790
<i>Mus musculus</i>	8 704 790
<i>Zea mays</i>	2 019 137
<i>Bos taurus</i>	1 669 337
<i>Sus scrofa</i>	1 559 495
<i>Arabidopsis thaliana</i>	1 529 700
<i>Danio rerio</i>	1 488 275
<i>Glycine max</i>	1 461 722
<i>Xenopus tropicalis</i>	1 271 480
<i>Oryza sativa</i>	1 253 557
<i>Ciona intestinalis</i>	1 205 674
<i>Triticum aestivum</i>	1 286 372
<i>Rattus norvegicus</i>	1 162 136

RNA (RNA totale, frazione poliadenilata, frazione a basso peso molecolare che comprende i miRNA ecc.), che possono o meno essere orientate in funzione del protocollo utilizzato.

L'analisi del profilo trascrizionale, dei geni condotto fino a qualche tempo fa mediante l'analisi di sequenze EST e più recentemente di dati RNA-Seq, consiste nel determinare quali trascritti sono espressi e la loro abbondanza relativa. In questo modo è possibile caratterizzare il piano di espressione specifico di diversi tipi cellulari (o compartimenti subcellulari), in condizioni fisiologiche o patologiche. È stato così possibile osservare, per esempio, che le cellule tumorali hanno un profilo di trascrizione sensibilmente diverso dalle cellule normali dello stesso tipo cellulare, e identificare biomarcatori trascrizionali condizione-specifici e quindi dotati di capacità diagnostica o prognostica.

Tuttavia, l'analisi accurata del profilo trascrizionale dei geni è molto laboriosa e complessa sia perché le letture di sequenza (*read*) normalmente coprono solo una porzione relativamente piccola del trascritto completo (*full-length*) sia per il fatto che uno stesso gene può esprimere un numero anche molto grande di trascritti (e quindi proteine) differenti a causa di iniziatori e terminatori alternativi della trascrizione e di eventi di splicing alternativo. Per questo non è banale assegnare una specifica lettura di sequenza a un determinato gene/trascritto e conseguentemente anche ottenere stime quantitative accurate dei livelli di espressione dei diversi trascritti alternativi. Al fine di determinare nel modo più accurato possibile il profilo trascrizionale dei geni a livello qualitativo e quantitativo sono stati messi a punto protocolli sperimentali sempre più efficienti per la preparazione delle librerie e sistemi di analisi bioinformatica sempre più sofisticati. Sono stati anche sviluppati protocolli sperimentali per il sequenziamento di specifiche frazioni e porzioni dell'RNA. Per esempio, è possibile preparare la libreria di cDNA in modo appropriato per garantire che l'intero mRNA, dalla coda di poli(A) al "cap", venga retrotrascritto in cDNA. A questo scopo può essere adottato il **protocollo dell'oligo-capping** o CAGE-Seq, in cui un oligonucleotide a sequenza nota viene aggiunto selettivamente agli mRNA dotati di cap, e quindi completi al 5'. In questo modo il cDNA completo può essere ottenuto utilizzando per la

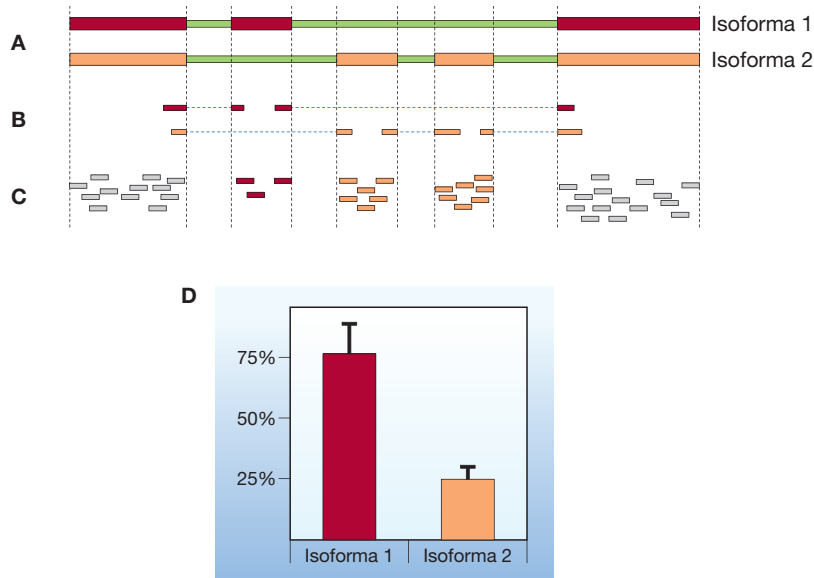


Figura 21.6 Identificazione, mediante analisi di dati RNA-Seq, dei due mRNA full-length espressi da uno stesso gene e determinazione della loro abbondanza relativa. (A) Le due varianti di splicing (isoforme 1 e 2) dello stesso trascritto primario a cui possono essere allineate le read ottenute dal sequenziamento. (B) Alcune delle read contengono i siti di splicing e ci informano sulla connettività degli esoni. (C) Delle rimanenti read, che corrispondono a regioni di singoli esoni, alcune (colorate in grigio) non sono in grado di discriminare tra i due trascritti alternativi in quanto si allineano ugualmente entrambi, mentre altre (colorate in rosso e in arancio) possono essere univocamente associate alle due diverse varianti di splicing. (D) Sulla base dei conteggi relativi delle diverse tipologie di read è possibile ottenere una determinazione sufficientemente accurata dei livelli di espressione dei due trascritti alternativi. Naturalmente, quanto più complesso è il pattern di splicing, tanto meno affidabili saranno le stime ottenute.

sintesi del primo filamento un oligo-dT ancorato, che si appaia selettivamente alla giunzione della coda di poli(A), mentre un primer complementare all'oligonucleotide adattatore al 5' viene utilizzato per la sintesi del secondo filamento. Attraverso la tecnica detta PET-Seq (*Paired-End Tag sequencing*) o SMORE-Seq è possibile ottenere letture di sequenza appaiate di entrambe le estremità dei trascritti espressi. Accoppiando dati PET-Seq e RNA-Seq tradizionali è possibile, mediante l'uso di appropriati sistemi di analisi bioinformatica, ottenere determinazioni sufficientemente attendibili dei diversi trascritti espressi e la loro abbondanza relativa (fig. 21.6).

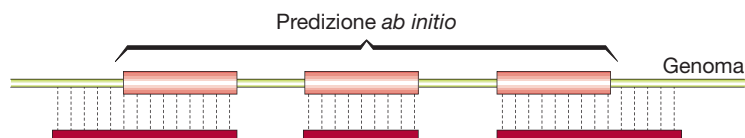
Nel caso di trascritti molto lunghi, non presenti nelle banche dati, e con un pattern di splicing alternativo complesso, e particolarmente quando non sia disponibile la sequenza genomica corrispondente, i dati RNA-Seq possono non essere in grado di effettuare una ricostruzione accurata dei trascritti completi e quindi è necessario procedere al clonaggio e al sequenziamento completo degli inserti di cDNA (fig. 21.5). Tuttavia, la possibilità di ottenere sequenze molto più lunghe dai sistemi di sequenziamento di terza generazione sta rapidamente risolvendo anche queste problematiche. Infine, sono in fase avanzate di studio alcune metodologie per il sequenziamento diretto dell'RNA evitando la laboriosa fase della sua trasformazione in cDNA. Queste metodologie potranno anche facilitare moltissimo lo studio dell'RNA editing (*vedi par. 13.2*) che sta assumendo un ruolo sempre più rilevante in numerosi pathway di regolazione dell'espressione genica.

Il sequenziamento del trascrittoma può anche consentirci di verificare la presenza di trascritti di fusione derivanti da geni diversi (ad es. il trascritto *bcr-abl* che si osserva nella leucemia mieloide cronica a seguito di una traslocazione cromosomica).

21.3 Identificazione e annotazione di geni

Una volta completato un progetto di sequenziamento genomico, la prima analisi che viene solitamente condotta riguarda l'identificazione e l'annotazione dei geni

Figura 21.7 Sistemi di predizione genica *ab initio*. Questi predicono esclusivamente le porzioni codificanti del gene (in rosso chiaro). La predizione può essere successivamente validata attraverso l'allineamento del trascritto corrispondente identificato nel sequenziamento del trascrittoma (in rosso scuro). L'evidenza trascrizionale consente anche di identificare le regioni 5' e 3' non tradotte, a monte e a valle della regione codificante.



codificanti proteine. Questa analisi è relativamente semplice nel caso di genomi procariotici, in quanto i geni sono privi di introni. Pertanto, la maggior parte dei geni può essere determinata semplicemente identificando le **Open Reading Frame (ORF)**, ovvero quei tratti genomici costituiti da un significativo numero di triplette (codoni) delimitati da un codone di inizio (AUG) e da un codone di stop (UAA, UAG o UGA). Una ORF è probabilmente corrispondente alla reale regione codificante di un gene (*Coding Sequence*, CDS) se ha una lunghezza significativamente maggiore rispetto a quella attesa in base alle proprietà del codice genetico (*vedi* cap. 3). In pratica, se consideriamo che nel codice genetico le triplette “senso” sono 61 e quelle “non senso” sono 3, la lunghezza attesa di una ORF dovrebbe essere di circa 20 triplette. Se tale lunghezza risulta significativamente maggiore (ad es. >100), sarà molto probabile che essa corrisponda a una reale regione codificante. Rimane il problema d'identificare geni codificanti per proteine di piccole dimensioni. A questo scopo sono stati sviluppati programmi bioinformatici sofisticati che tengono conto della composizione in basi delle ORF, che, normalmente, risulta diversa nel caso di ORF realmente codificanti rispetto a ORF prive di funzione o di analisi comparative con ORF omologhe in altri organismi (ad es. le sostituzioni di basi che si osservano nelle ORF corrispondono prevalentemente a cambiamenti “sinonimi” che non alterano l'amminoacido codificato dal codone, *vedi* par. 21.8).

Il problema si complica per i geni eucariotici, che possono essere interrotti da introni. In questo caso, l'identificazione e annotazione di geni in un genoma eucariotico neosequenziato si ottiene innanzitutto mediante analisi di genomica comparata, attraverso il confronto con un altro genoma nel quale i geni siano stati già annotati. Va sottolineato che, come abbiamo già detto in precedenza, il progetto di sequenziamento di un nuovo genoma dovrebbe sempre includere in parallelo anche il sequenziamento del trascrittoma. Le informazioni derivanti dall'analisi del trascrittoma sono fondamentali per ottenere una corretta annotazione dei geni, inclusa la determinazione dei siti alternativi di inizio e fine della trascrizione e di eventi di splicing alternativo. Difatti, la mappatura di trascritti completi o parziali sul genoma, utilizzando appropriati programmi di allineamento (*vedi* oltre), serve a delimitare in modo preciso le regioni esoniche dei geni, incluse le regioni non tradotte al 5' e al 3' che non vengono normalmente identificate dall'analisi comparata. Nel caso in cui non siano disponibili dati trascrizionali è possibile utilizzare programmi bioinformatici più complessi che effettuano la predizione genica *ab initio*, basati sia sull'analisi delle proprietà composizionali delle sequenze codificanti che sulla presenza di motivi oligonucleotidici tipicamente localizzati nei geni (codone d'inizio, siti di splicing donatore e accettore, codone di stop, segnale di poliadenilazione ecc.) (**fig. 21.7**).

21.4 Confronto e allineamento di sequenze

Il confronto e l'allineamento di sequenze sono di gran lunga le applicazioni bioinformatiche più diffuse. Infatti, se abbiamo tra le mani una sequenza che desideriamo caratterizzare, sia essa nucleotidica o amminoacidica, l'analisi più semplice

consiste nel confrontarla con tutte le altre sequenze note. In altre parole, ci chiediamo se vi siano una o più sequenze note che mostrino un significativo grado di somiglianza con la sequenza in esame. Questo problema bioinformatico viene affrontato applicando dei programmi in grado di costruire allineamenti "locali" tra la sequenza data e tutte quelle presenti nella banca dati. Il programma più utilizzato per questo tipo di analisi, ideato da David J. Lipman e collaboratori (fig. 21.8), è denominato **BLAST** (*Basic Local Alignment Search Tool*), il cui risultato consiste in un elenco di sequenze simili in ordine decrescente di significatività statistica (fig. 21.9).

Il confronto può essere effettuato a livello nucleotidico (DNA o RNA) o a livello amminoacidico (proteine). Il confronto a livello delle proteine è preferibile, se possibile, dato che risulta molto più selettivo. Infatti, data la diversa natura dell'alfabeto delle basi rispetto a quello degli amminoacidi (4 caratteri contro 20), gli allineamenti casuali tra sequenze proteiche sono molto meno probabili che tra sequenze nucleotidiche. È possibile fare il confronto a livello proteico anche partendo da sequenze nucleotidiche: in questo caso sia la sequenza in esame che tutte le sequenze presenti nella banca dati vengono tradotte in ipotetiche sequenze proteiche utilizzando tutte le sei possibili fasi di lettura (tre su ciascuno dei due filamenti di DNA).

Mentre l'allineamento locale cerca di identificare se due sequenze contengono un tratto comune significativamente simile, l'allineamento globale cerca di allineare tutti i residui di una sequenza con tutti quelli di un'altra sequenza. Se costruiamo un allineamento globale, infatti, assumiamo che le due sequenze siano omologhe, ovvero abbiano un progenitore comune dal quale hanno subito divergenza in seguito al processo evolutivo. Se le sequenze omologhe di cui disponiamo sono più di due, possiamo costruire, attraverso diversi programmi bioinformatici come **CLUSTAL** o **MUSCLE**, un allineamento multiplo che, evidenziando i siti maggiormente conservati o variabili, risulterà molto utile per procedere alla caratterizzazione funzionale e strutturale della proteina con approcci sperimentali. La costruzione di un allineamento multiplo (fig. 21.10) fornisce indicazioni sul livello complessivo di conservazione di una proteina, evidenziando sia i siti conservati che quelli variabili.



Figura 21.8 David J. Lipman. Lipman è uno degli ideatori dell'algoritmo BLAST ed è attualmente direttore dell'NCBI (National Center for Biotechnology Information, USA).

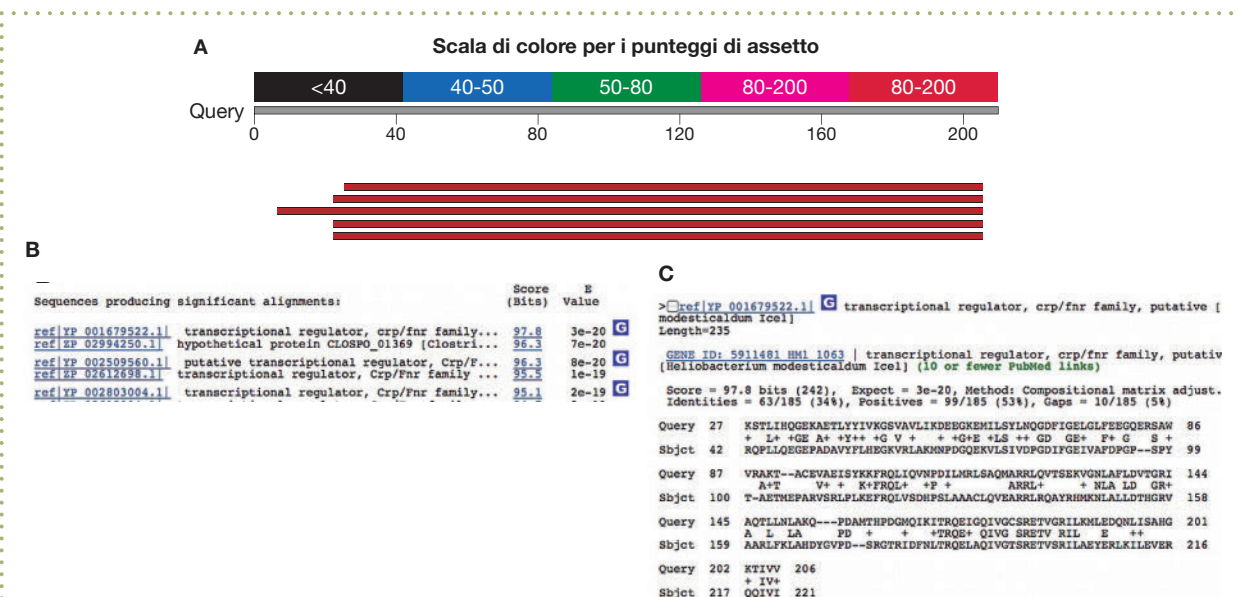


Figura 21.9 Esempio di output generato dal programma BLAST. Esito del confronto fra la sequenza della proteina CAP di *Escherichia coli* (accession number: P0ACJ8, 210 aa) e tutte le proteine note del raggruppamento tassonomico dei clostridi. (A) Rappresentazione grafica dell'allineamento in cui le sequenze che si allineano con la proteina sonda (query) sono rappresentate da linee in colore corrispondente al punteggio dell'allineamento. (B) Elenco delle sequenze risultate più simili in ordine decrescente di similarità. (C) Allineamento che mostra i residui identici e quelli simili (+) tra la sequenza query e una delle sequenze della banca dati.

neano con la proteina sonda (query) sono rappresentate da linee in colore corrispondente al punteggio dell'allineamento. (B) Elenco delle sequenze risultate più simili in ordine decrescente di similarità. (C) Allineamento che mostra i residui identici e quelli simili (+) tra la sequenza query e una delle sequenze della banca dati.

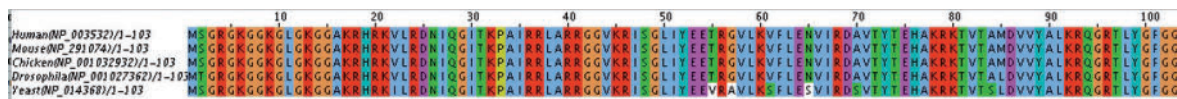


Figura 21.10 Allineamento multiplo dell'istone H4 di uomo, topo, pollo, *Drosophila* e lievito. Gli amminoacidi sono indicati secondo il codice IUB a una lettera.

L'allineamento tra sequenze è l'approccio bioinformatico più efficace per la caratterizzazione funzionale delle sequenze. Per esempio, noto il corredo di geni, più o meno completo, presente in un genoma, la strategia bioinformatica più semplice e veloce per l'identificazione della loro funzione prevede il confronto delle ipotetiche sequenze proteiche con l'intera collezione delle proteine a funzione nota. Tale confronto viene normalmente condotto con programmi di allineamento locale, molto efficienti dal punto di vista computazionale, come BLAST, già citato in precedenza. L'osservazione di un grado di similarità statisticamente significativo con una o più proteine a funzione nota consente, per analogia, di attribuire la stessa funzione anche a proteine ignote. È evidente che tale funzione andrà poi verificata con approcci sperimentali appropriati, ma l'analisi bioinformatica ci avrà aiutato a circoscrivere in modo significativo lo spettro delle ipotesi da verificare. Per l'annotazione delle sequenze proteiche possono anche essere utilizzati programmi in grado d'identificare specifiche regioni della proteina a funzione nota (domini in grado di legare DNA o RNA, segmenti trans-membrana, peptidi segnale che indirizzano la proteina verso specifici compartimenti cellulari ecc.) che possono contribuire in modo significativo alla sua caratterizzazione funzionale.

21.5 Identificazione e annotazione di regioni regolatorie e caratterizzazione delle proprietà epigenetiche della cromatina

Abbiamo visto che l'espressione dei geni è un processo finemente regolato secondo un meccanismo che comporta interazioni specifiche tra proteine e DNA o RNA. Per esempio, l'RNA polimerasi che deve iniziare la trascrizione a siti specifici non si lega a caso sul DNA, ma riconosce particolari sequenze a monte dei geni (*vedi* capp. 9-11). Nei batteri la specificità dell'interazione tra l'RNA polimerasi e il promotore viene garantita dal fattore σ che riconosce brevi sequenze del promotore in posizione -35 e -10 , rispetto al sito di inizio della trascrizione. La specificità del riconoscimento è molto più complessa negli eucarioti e coinvolge un gran numero di fattori di trascrizione che riconoscono siti di legame sul DNA localizzati sia nella regione del promotore che in altre regioni anche molto distanti, a monte e a valle di questo. Questi siti di legame sono normalmente brevi (da 6 a 15 pb) e degenerati, ovvero solo poche posizioni sono evolutivamente conservate e diverse varianti sono ugualmente attive nel legame, anche se con una differente affinità.

Tipicamente questi siti sono rappresentati da sequenze consenso, in cui per ciascuna posizione viene riportata la base (o le basi, utilizzando il codice IUB per le basi ambigue; **tab. 21.2**) più rappresentata. Per esempio, confrontando 250 siti di legame corrispondenti all'elemento -35 di promotori noti del batterio *Escherichia coli*, di lunghezza pari a 6 pb, si osservano ben 147 varianti delle 1024 possibili, la cui sequenza consenso è TTGACA.

Una modalità certamente più intuitiva e ugualmente informativa per rappresentare questi motivi è il **sequence logo**, che fornisce una rappresentazione grafica del motivo in cui l'altezza di ciascuna base è proporzionale al suo grado di conservazione (**finestra 21.1**).

Tabella 21.2 A Codice IUB per le basi ambigue.

Simbolo	Significato	Origine di destinazione
G	G	Guanina
A	A	Adenina
T	T	Timina
C	C	Citosina
R	G o A	puRina
Y	T o C	pYrimidina
M	A o C	aMino
K	G o T	Keto
S	G o C	Interazione forte (3 legami H)
W	A o T	Interazione debole (2 legami H)
H	A o C o T	non-G, H segue G nell'alfabeto
B	G o T o C	non-A, B segue A
V	G o C o A	non-T (non-U), V segue U
D	G o A o T	non-C, D segue C
N	G o A o T o C	aNy

Tabella 21.2 B Codici IUB per gli amminoacidi.

Amminoacido	Codice IUB (3 lettere)	Codice IUB (1 lettera)
Alanina	Ala	A
Arginina	Arg	R
Asparagina	Asn	N
Acido aspartico	Asp	D
Cisteina	Cys	C
Acido glutammico	Glu	E
Glutamina	Gln	Q
Glicina	Gly	G
Istidina	His	H
Isoleucina	Ile	I
Leucina	Leu	L
Lisina	Lys	K
Metionina	Met	M
Fenilalanina	Phe	F
Prolina	Pro	P
Serina	Ser	S
Treonina	Thr	T
Triptofano	Trp	W
Tirosina	Tyr	Y
Valina	Val	V

FINESTRA 21.1 RAPPRESENTAZIONE DI MOTIVI REGOLATIVI MEDIANTE IL SEQUENCE LOGO

Dato un insieme di motivi regolativi sperimentalmente validati, possiamo costruire una matrice posizionale delle frequenze (PWM, *Position Weight Matrix*) costituita da 4 righe, una per ciascuna base, e un numero di colonne pari alla lunghezza del motivo. Una cella di questa matrice riporta la frequenza di una determinata base nella posizione corrispondente. Per esempio, se analizziamo gli elementi -35 di 250 promotori di *Escherichia coli* (arep.med.harvard.edu/ecoli_matrices/rpoD.html) si ottiene la PWM sotto rappresentata:

pos	1	2	3	4	5	6
A	16	18	21	137	55	143
C	19	10	24	46	126	24
G	13	12	161	20	20	41
T	202	210	44	47	49	42

Dalla PWM possiamo calcolare un indice dell'entropia di ciascuna posizione $H(i)$ che sarà massima se le basi sono usate con la stessa frequenza:

$$H(i) = - \sum_{b=A,C,G,T} f(b,i) \cdot \log_2 f(b,i)$$

dove $f(b,i)$ corrisponde alla frequenza della base b nella posizione i (che in questo caso varia da 1 a 6). L'indice di entropia

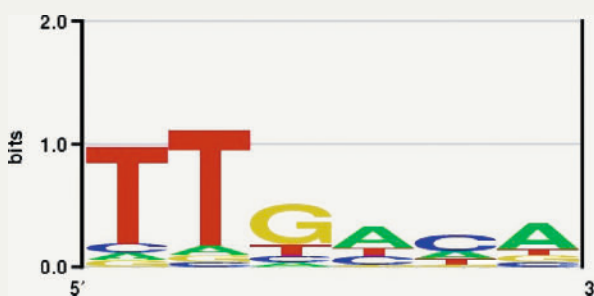
nella posizione i è inversamente correlato alla quantità di informazione $I(i)$, che è data da:

$$I(i) = 2 - H(i)$$

Noto $I(i)$ per ciascuna posizione, possiamo calcolare l'altezza di ciascuna base $h(i,b)$ in modo proporzionale alla sua frequenza:

$$h(i,b) = I(i) \cdot f(b,i)$$

Possiamo così ottenere la rappresentazione grafica del motivo come quella di seguito rappresentata. Si noti che anche se la sequenza consenso è TTGACA, la visualizzazione grafica attraverso il logo ci consente di apprezzare che solo le prime due posizioni del motivo sono conservate in modo significativo.



Per identificare e riconoscere questi elementi regolativi, in particolare i siti di legame di fattori di trascrizione, possono essere utilizzati dei particolari programmi bioinformatici, detti di **pattern discovery**, che sono in grado di individuare motivi oligonucleotidici sovrarappresentati nell'insieme delle sequenze in esame. La costruzione dell'insieme di sequenze da analizzare è un elemento critico per l'attendibilità dell'analisi.

Anche lo studio degli elementi regolativi ha ricevuto un eccezionale impulso dall'avvento delle piattaforme NGS. Gli esperimenti di **ChIP-Seq** (vedi par. 20.12 e fig. 21.11) consentono di avere una mappa su scala genomica delle interazioni tra specifiche proteine e il DNA. In questo modo è possibile identificare nel genoma tutti i siti di legame di un determinato fattore di trascrizione o anche ottenere una mappatura completa di specifiche modificazioni istoniche. L'analisi bioinformatica dei dati di sequenziamento derivati dagli esperimenti di ChIP-Seq consiste nel mappaggio sul genoma delle read ottenute, che, come abbiamo già visto nel paragrafo 20.12, provengono prevalentemente dalle regioni di interazione con la proteina immunoprecipitata, in modo da identificare "picchi" (per i fattori di trascrizione) o "altopiani" (per le modificazioni istoniche) che delimitano la regione di interazione (fig. 21.11).

Analogamente, gli approcci NGS consentono anche di condurre analisi su larga scala del pattern di metilazione del DNA genomico (**studio del metiloma**). Questo è possibile sia mediante esperimenti di ChIP-Seq, utilizzando anticorpi specifici per le citosine metilate, sia sequenziando il DNA prima e dopo il trattamento con bisolfito, che converte in modo selettivo le citosine non metilate in uracile (vedi finestra 17.2). L'allineamento sul genoma di riferimento delle read

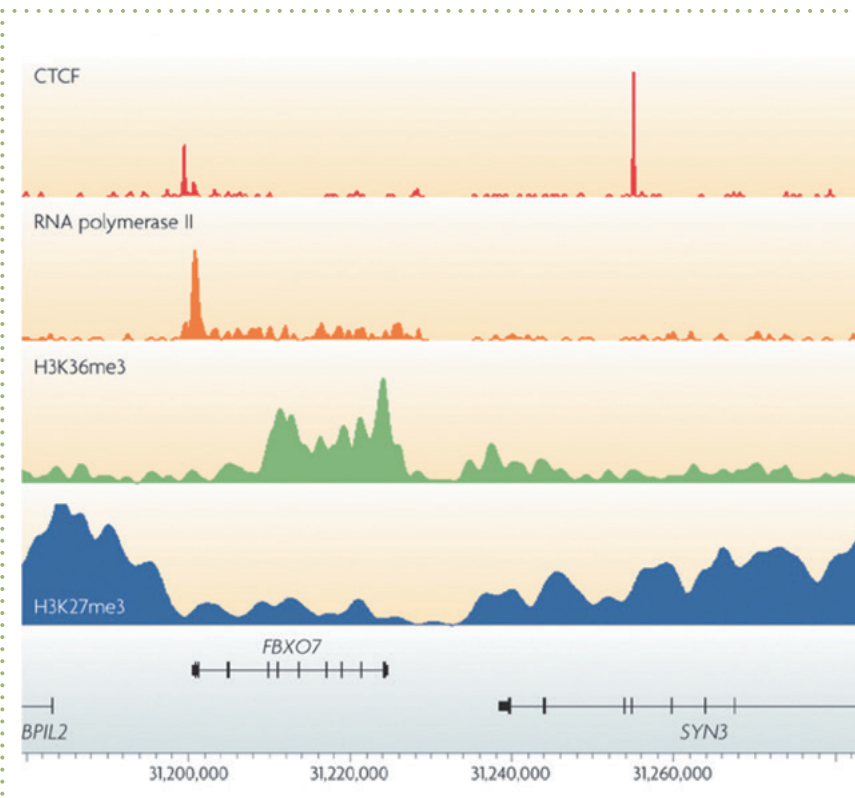


Figura 21.11 Determinazione delle interazioni DNA-proteine mediante esperimenti di ChIP-Seq. Il mappaggio delle read ottenuto dai frammenti di DNA immunoprecipitati individua dei picchi o altipiani (in ordinata si riporta il numero di read mappate in funzione della posizione lungo il cromosoma 22 riportata in ascissa). Nell'esempio mostrato che si riferisce a cellule T umane si osservano picchi per l'interazione della proteina CTCF che si lega alle regioni "insulator" (rosso) e per RNA polimerasi II (arancio). Regioni di arricchimento più estese si osservano per le modificazioni istoniche come H3K36me3 (verde), normalmente associata a regioni trascrizionalmente attive, e H3K27me3 (blu), che invece è associata a regioni repressate trascrizionalmente ad opera del complesso Polycomb. Si noti come è possibile interpretare il profilo epigenetico mostrato con l'espressione del gene *FBXO7* e il silenziamento del gene *SYN3*. [Fonte: Park PJ. (2009), "ChIP-seq: advantages and challenges of a maturing technology", *Nat Rev Genet*, 10(10), pp. 669-80, doi: 10.1038/nrg2641.]

ottenute da DNA trattato e non trattato con bisolfito identifica con precisione i siti metilati.

Integrando i risultati di diversi esperimenti condotti sugli stessi campioni (ChIP-Seq di fattori di trascrizione e varie modificazioni istoniche, analisi del metiloma, analisi del trascrittoma) è possibile ottenere una mappa dettagliata delle proprietà epigenetiche della cromatina e della dinamica del profilo di espressione in diverse condizioni e tipi cellulari.

21.6 Annotazione delle sequenze proteiche

Anche se la ricerca di similarità non ha evidenziato alcuna proteina significativamente simile, e quindi realisticamente omologa a quella in esame, è possibile utilizzare approcci bioinformatici alternativi in grado di effettuare una serie di predizioni sulle caratteristiche strutturali e funzionali delle stesse. In particolare, attraverso l'applicazione di questi programmi è possibile identificare domini proteici noti (ad es. *zinc finger*) che possono fornire indicazioni utili alla caratterizzazione della funzione di proteine. I programmi di predizione possono indicare, con un buon livello di affidabilità, molte altre caratteristiche, come:

- peptidi segnale, che indirizzano la proteina verso specifici compartimenti cellulari (ad es. mitocondri);
- segmenti trans-membrana;
- elementi di struttura secondaria (α -elica o foglietto- β);
- siti di modificazione post-traduzionale (ad es. fosforilazione, acetilazione, glicosilazione ecc.).

I risultati di queste predizioni possono indirizzare le successive validazioni sperimentali volte alla completa caratterizzazione biochimica, funzionale e strutturale delle proteine.

21.7 Banche dati e browser genomici

Non appena furono prodotte le prime sequenze nucleotidiche e proteiche, ci si rese conto della necessità di sistemi informatizzati adeguati per il loro immagazzinamento, gestione e analisi. Nacquero così le banche dati biomolecolari. La prima fu l'**EMBL data library**, istituita nel 1980 dal Laboratorio Europeo di Biologia Molecolare (EMBL), cui seguirono la **Genbank** negli Stati Uniti e la **DDBJ** in Giappone. Possiamo consultare indifferentemente le banche dati EMBL, Genbank e DDBJ in quanto esse contengono gli stessi dati e le stesse informazioni. Infatti, a seguito di un accordo internazionale vengono "sincronizzate". La ragione principale per cui le banche dati sono mantenute in triplicato è quella di garantire il massimo livello di sicurezza alla preservazione dei dati, evitando il rischio che tutta l'enorme quantità di informazioni immagazzinate possa andare persa per una eventuale catastrofe locale.

Ciascuna sequenza contenuta nella banca dati costituisce una "entry" che, oltre alla sequenza, contiene una serie d'informazioni supplementari, quali la specie, la classificazione tassonomica, le coordinate di tutti gli elementi funzionali in essa contenuti (gene, esone, introne ecc.).

Quando un gruppo di ricerca determina una sequenza nucleotidica, questa, per essere utilizzata in una pubblicazione, deve essere depositata nella banca dati per poi essere resa pubblica. La procedura di sottomissione alla banca dati prevede che a ciascuna sequenza venga assegnato un codice d'identificazione univoco, detto **accession number**, che viene mantenuto nel tempo anche se la entry è sottoposta a successivi aggiornamenti. Parallelamente sono stati sviluppati sistemi d'interrogazione delle banche dati per consentire un'agevole e veloce ricerca delle informazioni in esse contenute e la selezione e l'estrazione di specifici sottoinsiemi di sequenze.

Oltre alle banche dati che contengono le sequenze nucleotidiche e proteiche, definite banche dati primarie, vi è un gran numero di banche dati specializzate che collezionano particolari sottoinsiemi omogenei di dati, corredati di informazioni addizionali, anche di natura sperimentale (ad es. *MirBase* è la banca dati che colleziona tutti i miRNA noti nell'uomo e in altri organismi). Attualmente, il sistema di banche dati più utilizzato è quello allestito presso il centro di bioinformatica degli USA (**National Center for Biotechnology Information, NCBI**). Vi sono molti altri centri di bioinformatica di livello internazionale, come l'**EBI (European Bioinformatics Institute)**, che, come l'NCBI, mettono a disposizione dei ricercatori collezioni di banche dati e programmi bioinformatici.

Le banche dati primarie, sopra citate, non sono strutturate in modo adeguato per raccogliere tutte le informazioni legate ai progetti di sequenziamento dei genomi completi. Per questo scopo sono state sviluppate delle risorse bioinformatiche apposite, denominate **browser genomici**. Un browser genomico raccoglie tutte le informazioni relative a uno o più genomi, come l'annotazione dei geni e di altre caratteristiche strutturali e funzionali, quali la localizzazione di geni, esoni, introni, siti polimorfici ecc. Ciascuna annotazione elementare è caratterizzata da una precisa localizzazione sul genoma, rappresentata dall'indicazione del cromosoma e dalla posizione d'inizio e fine nello stesso. I browser genomici più diffusi, relativi ai genomi di metazoi, sono quelli allestiti presso l'EBI (**ENSEMBL**) e l'Università della California a Santa Cruz (**UCSC**). Dato che le coordinate genomiche assolute rimangono invariate finché non venga rilasciata una nuova versione del genoma, tutte le annotazioni ottenute applicando varie strategie possono essere incrociate tra loro in modo da formare una mappa più o meno dettagliata degli elementi strutturali e funzionali presenti in una qualsiasi regione del genoma (**fig. 21.12**). Oltre ai browser genomici più diffusi, sopra citati, ve ne sono molti altri specifici per determinati raggruppamenti tassonomici (batteri, funghi, piante ecc.).

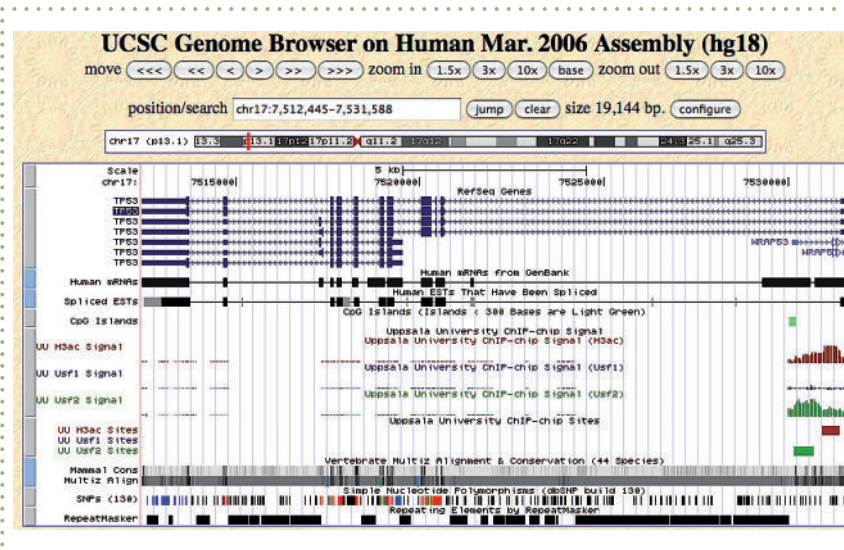


Figura 21.12 Esempio di visualizzazione nel browser UCSC della regione genomica relativa al gene *TP53* umano localizzato nel cromosoma 17.

21.8 Evoluzione molecolare

L'allineamento di geni omologhi, sia a livello di nucleotidi che di proteine, riflette il processo evolutivo che questi hanno subito in seguito alla loro divergenza. Il punto di divergenza coincide con un evento di speciazione, oppure con un evento di duplicazione genica. I geni omologhi che hanno cominciato a evolvere in modo indipendente in seguito a speciazione sono detti **ortologhi** (ad es. β -globina in uomo e topo), mentre quelli che si sono originati grazie a un evento di duplicazione genica sono detti **paraloghi** (ad es. α - e β -globina nell'uomo). Per definizione, tutti i geni membri di una stessa famiglia genica all'interno del genoma sono paraloghi. Gli studi di evoluzione molecolare si propongono di ricostruire la storia evolutiva dei geni, che, nel caso dei geni ortologhi, corrisponde con la storia evolutiva delle specie corrispondenti.

Il motore del processo evolutivo sono le mutazioni, che possono essere spontanee, dovute a errori intrinseci nel processo di replicazione, o indotte da fattori ambientali (ad es. esposizione a radiazioni o agenti chimici mutageni). La maggior parte delle mutazioni viene riconosciuta ed eliminata dai sistemi di riparazione (*vedi* cap. 7), ma una piccola frazione di queste, se a carico di cellule germinali, può essere trasmessa e fissata nella progenie, per diffondersi nell'intera popolazione. Le mutazioni possono essere di diversa natura, quali le sostituzioni di singole basi o inserzioni/delezioni/traslocazioni di regioni più o meno ampie del genoma. Gli studi di evoluzione molecolare sono, di norma, focalizzati sulle sostituzioni puntiformi. La maggior parte delle sostituzioni nucleotidiche puntiformi è evolutivamente "neutrale", in quanto non produce conseguenze evidenti sul fenotipo; sono, solitamente, definite come **polimorfismi** o anche **SNP** (*Single Nucleotide Polymorphisms*). Classicamente, si definisce SNP una sostituzione nucleotidica che abbia una frequenza allelica nella popolazione almeno pari all'1%. Il termine "mutazione" viene, invece, usato per le sostituzioni nucleotidiche più rare cui, normalmente, sono associati fenotipi patologici. L'analisi degli SNP è molto importante per vari aspetti, per esempio per applicazioni nel campo della farmacogenomica, e molti studi sono stati finora condotti per la loro analisi in diverse popolazioni umane. Attualmente, risultano caratterizzati nel genoma umano oltre 18 milioni di SNP, che sono stati depositati in una banca dati specializzata e visualizzabili nei browser genomici come UCSC. Le sostituzioni nucleotidiche possono essere presenti sia nelle regioni codificanti che non codificanti del genoma. Le **sostituzioni** nelle regioni codificanti

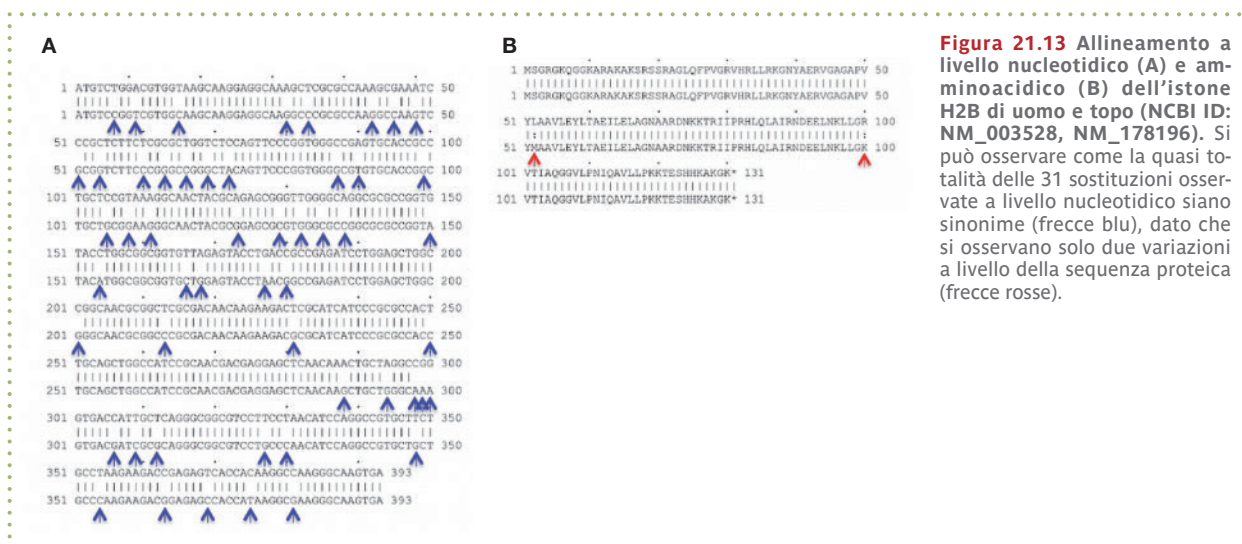


Figura 21.13 Allineamento a livello nucleotidico (A) e aminoacidico (B) dell'istone H2B di uomo e topo (NCBI ID: NM_003528, NM_178196). Si può osservare come la quasi totalità delle 31 sostituzioni osservate a livello nucleotidico siano sinonime (freccie blu), dato che si osservano solo due variazioni a livello della sequenza proteica (freccie rosse).

sono classificate come **non-sinonime** o **sinonime**, a seconda che producano o meno una alterazione nel significato del codone. Le sostituzioni sinonime, che non alterano la sequenza amminoacidica, non essendo sottoposte a particolari pressioni selettive, sono molto più frequenti delle sostituzioni non sinonime. La misura del tasso di sostituzioni nucleotidiche nel confronto di due o più geni omologhi deve essere, quindi, effettuata distinguendo i cambiamenti sinonimi da quelli non sinonimi. Il tasso delle sostituzioni sinonime, evolutivamente neutri, riflette la dinamica generale del processo mutazionale in un determinato genoma o regione genomica. Al contrario, il tasso di mutazione non sinonima, specifico per ciascun gene, riflette il livello delle costrizioni funzionali del suo prodotto d'espressione. Per esempio, le proteine istoniche, evolutivamente molto conservate, hanno un tasso di mutazione non sinonima molto più basso di quello delle mutazioni sinonime (fig. 21.13). Il tasso di mutazione può essere espresso in termini relativi, in termini di sostituzioni/sito o in termini assoluti, se è nota la divergenza temporale tra i geni in esame. Tipicamente, il tasso delle mutazioni non sinonime è circa 10^{-10} sostituzioni/(sito × anno), mentre quello delle mutazioni sinonime è di oltre un ordine di grandezza maggiore. Può essere utile anche analizzare il tasso delle sostituzioni nucleotidiche nelle regioni non codificanti, al fine di identificare specifici tratti conservati che potrebbero corrispondere a regioni regolative deputate al controllo dell'espressione genica (promotori, enhancer ecc.). Il tasso effettivo delle sostituzioni nucleotidiche può essere calcolato, sulla base delle sostituzioni osservate, applicando metodologie appropriate, normalmente basate su approcci di massima verosimiglianza (*maximum likelihood*).

Il grado di divergenza tra due geni è legato sia alla specifica dinamica evolutiva che al tempo di divergenza, ovvero il tempo intercorso a partire dal progenitore più recente dei geni omologhi in esame. La correlazione tra divergenza genetica e tempo di divergenza, in alcuni casi, può risultare più o meno lineare in funzione del tempo. Questa proprietà viene definita **orologio molecolare** e può essere molto utile per determinare la cronologia degli eventi evolutivi su scale temporali anche molto ampie (fig. 21.14). L'orologio molecolare può essere calibrato se sono noti i tempi di divergenza tra due o più gruppi tassonomici sulla base di evidenze paleontologiche.

Come illustrato in figura 21.14, le relazioni evolutive tra geni o specie possono essere rappresentate da un **albero filogenetico**, una sorta di grafo costituito da nodi e rami le cui estremità coincidono con le unità tassonomiche in esame,

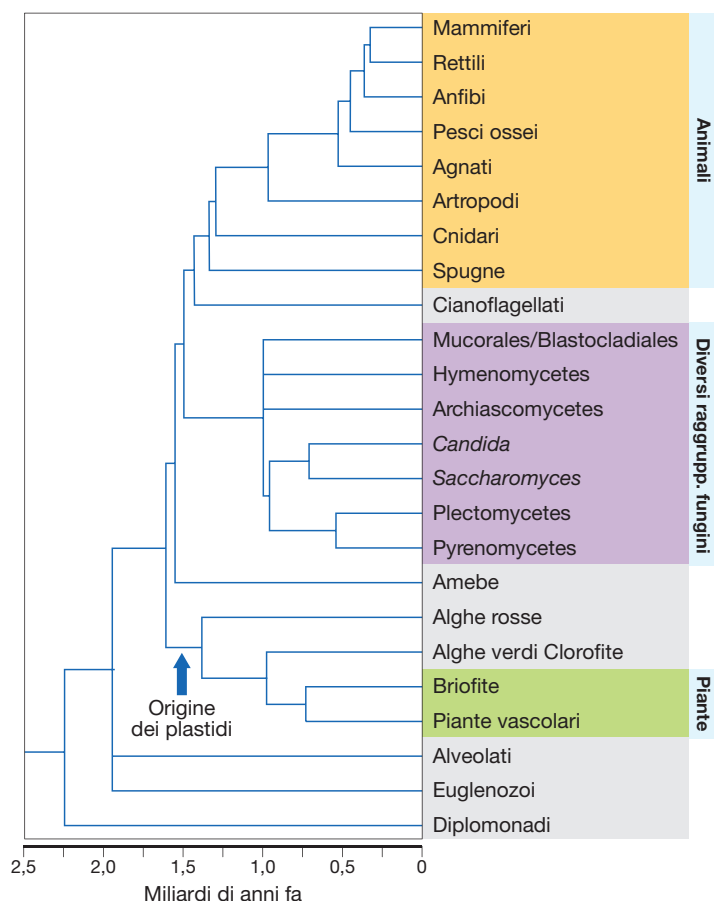


Figura 21.14 Scala temporale dell'evoluzione degli eucarioti. L'albero filogenetico rappresentato è basato sull'analisi evolutiva a livello amminoacidico, condotta con diversi strumenti bioinformatici, di un gran numero di proteine omologhe; per la calibrazione temporale è stata utilizzata la divergenza tra mammiferi e uccelli fissata a 310 milioni di anni fa. [Modificata da: Hedges *et al.*, (2004), *BMC Evolutionary Biology*, 4p.2 do-i:10.1186/1471-2148-4-2.]

congiunte da rami ai nodi interni che rappresentano le unità tassonomiche ancestrali. Gli alberi filogenetici possono essere orientati o non orientati, ovvero dotati o meno di una radice corrispondente al progenitore di tutte le unità tassonomiche considerate, e vengono determinati attraverso l'applicazione di sofisticate metodologie matematico-statistiche.

Indirizzi web delle principali risorse bioinformatiche

1. Il sito <http://www.genomesonline.org/gold.cgi>, aggiornato continuamente, riporta tutti i genomi completamente sequenziati.
2. Il sito <http://www.ncbi.nlm.nih.gov> del National Center of Biotechnology Information (NCBI) contiene un gran numero di banche dati e programmi di analisi di largo utilizzo.
3. Il browser genomico dell'Università della California a Santa Cruz (UCSC) è accessibile presso il sito <http://genome.ucsc.edu/>.
4. Il sito EMBOSS (<http://www.ebi.ac.uk/tools/emboss/>) raccoglie numerosi programmi di bioinformatica.

22.1 Il lievito *Saccharomyces cerevisiae*

22.2 Il moscerino della frutta *Drosophila melanogaster*

22.3 Il nematode *Caenorhabditis elegans*

22.4 L'anfibio *Xenopus laevis* e il suo successore *Xenopus tropicalis*

22.5 Il pesce *Danio rerio* o "zebrafish"

22.6 Il topo comune *Mus musculus*

22.7 La pianta *Arabidopsis thaliana*

A fronte delle decine di milioni di specie di organismi viventi presenti sulla Terra, lo studio di poche decine di "organismi modello" sta alla base di quasi tutto quello che sappiamo sui processi e meccanismi della vita: come la cellula cresce e si divide, come esprime la sua informazione genetica, come utilizza e immagazzina energia ecc. Il presupposto che giustifica l'utilizzo di organismi modello in Biologia è l'evidenza che i meccanismi molecolari di base dei processi biologici fondamentali sono stati estremamente conservati nel corso dell'evoluzione: ne deriva che quello che è vero per un batterio, o un lievito, è verosimilmente vero per un nematode, un topo e anche per l'uomo. Ovviamente esistono delle differenze fra questi organismi, ma i meccanismi molecolari di base sono molto simili.

Il termine "modello" viene usato spesso con diverse accezioni. In particolare, in questi ultimi anni si parla spesso di **modello animale**, con cui si intende, in genere, un topo il cui patrimonio genetico è stato modificato in modo da mimare una malattia genetica umana. In questo capitolo, invece, con il termine **organismi modello** si intendono organismi anche molto diversi fra di loro, ma accomunati dal fatto che sono stati particolarmente utili per comprendere i meccanismi molecolari di processi biologici complessi. Studiando un certo processo molecolare in lievito, per esempio, si possono ottenere risposte su come un carattere è controllato a livello molecolare e su come viene trasmesso di generazione in generazione. Le conclusioni ottenute su questo eucariote unicellulare sono verosimilmente valide anche per eucarioti pluricellulari, fenotipicamente molto più complessi. Tale assunzione è stata, di fatto, confermata tutte le volte che è stata verificata sperimentalmente. È per questo che la ricerca biologica fatta anche su organismi modello distanti dall'uomo, per esempio lievito o *Drosophila*, permette spesso di fare scoperte molto significative su processi molecolari di base il cui cattivo funzionamento nella specie umana determina spesso l'insorgenza di malattie.

Mentre molti processi biologici sono simili nella maggior parte degli organismi, spesso è molto più facile studiare alcuni particolari aspetti su un organismo piuttosto che su un altro. Per esempio, è più facile fare della genetica su piccoli organismi che si riproducono molto rapidamente e fare della biochimica su gran-

di animali da cui si possono ottenere notevoli quantità di tessuti. Da qui la scelta dell'organismo più opportuno per vari campi di ricerca biologica. Forse il primo ad avere scelto consapevolmente un organismo modello per le sue ricerche è stato Gregor Mendel con i suoi esperimenti di impollinazione incrociata tra diverse linee pure della pianta di pisello. Come ben sappiamo, i risultati da lui ottenuti su questo organismo modello sono alla base delle nostre conoscenze di Genetica valide per tutti gli organismi viventi. Un altro esempio è il batterio *Escherichia coli*, che è stato utilizzato per gli studi fondamentali sul codice genetico e sulla replicazione, trascrizione e traduzione, con risultati sostanzialmente validi anche per gli organismi superiori.

Come abbiamo detto, gli organismi modello utilizzati dai biologi non sono molti; inoltre, alcuni in auge un secolo fa non sono quasi più usati oggi, mentre nuovi modelli vengono ogni tanto proposti. I piselli sono stati presto abbandonati dalla Genetica, che si è rivolta ad altri organismi più facilmente studiabili: inizialmente soprattutto *Drosophila*, ma anche microrganismi eucariotici, prima *Neurospora* e poi i lieviti, o procariotici quali *E. coli* e i suoi virus (batteriofagi). Lo studio dell'embriologia e dello sviluppo animale ha utilizzato come organismi modello gli anfibi (dapprima rana e tritone, poi *Xenopus*), il riccio di mare e il pollo, ma anche organismi più semplici quale l'ameba *Dictyostelium*. I biochimici, interessati a studiare le proteine e altri componenti cellulari, avevano bisogno di grandi quantità di materiale e per molto tempo il materiale preferito erano tessuti e organi di grandi mammiferi, come per esempio fegato di bue o di maiale; oggi le nuove tecniche "miniaturizzate" permettono di lavorare su animali più piccoli e quindi il topo diventa l'organismo modello preferito. Va anche detto che ci sono stati e ci sono particolari organismi modello, usati da un numero limitato di ricercatori, che permettono di studiare specifici aspetti della Biologia. Alcuni esempi sono:

- il mais, che ha permesso alla genetista Barbara McClintock di scoprire e studiare i trasposoni;
- i molluschi calamaro e *Aplysia*, che rappresentano modelli sperimentali molto validi per studi di elettrofisiologia e neurobiologia;
- il pesce palla abissale *Fugu rubripes*, un vertebrato con genoma estremamente ridotto, circa 1/8 del genoma umano, molto utile negli studi di genomica comparata.

La scelta di organismi modello oggi viene fatta con molta più consapevolezza e oculatezza rispetto a cento anni fa. Nella prima metà del secolo scorso venivano scelti diversi organismi modello per le varie discipline biologiche (Genetica, Embriologia ecc.). In questi ultimi quarant'anni, invece, considerando l'unificazione che c'è stata tra le diverse discipline biologiche, si cerca di individuare organismi modello che si prestino bene ai diversi approcci sperimentali, genetico, biochimico, molecolare, cellulare ecc. Sono stati così introdotti con molto successo *Caenorhabditis elegans*, zebrafish e la pianta *Arabidopsis*; in tutti questi casi si è fatta molta attenzione a scegliere organismi che si prestano bene sia all'analisi genetica che allo studio dello sviluppo embrionale da ambedue i punti di vista, morfologico e molecolare.

In genere le principali caratteristiche che un organismo deve possedere per essere considerato un buon organismo modello sono le seguenti:

- il suo utilizzo deve essere economico;
- deve poter essere tenuto in condizioni controllate di laboratorio occupando il minor spazio possibile;
- deve avere un ciclo di riproduzione rapido;
- deve originare una progenie numerosa;
- la sequenza del suo genoma deve essere conosciuta;

- deve avere caratteristiche che permettano l'applicazione delle più moderne tecnologie genetiche e molecolari.

Descriveremo qui appresso solo alcuni degli organismi modello importanti per la Biologia in genere e per la Biologia Molecolare in particolare, tralasciando di parlare di *E. coli* e dei suoi batteriofagi, il cui ruolo come organismo modello è ampiamente sottolineato nel testo. Ribadiamo anche il fatto che oltre a quelli qui descritti esistono altri organismi che sono usati come modelli per specifici aspetti della Biologia o come modelli dei diversi gruppi filogenetici.

22.1 Il lievito *Saccharomyces cerevisiae*

Il lievito di birra *Saccharomyces cerevisiae* è uno dei sistemi modello più utilizzati per tutta una serie di caratteristiche che descriveremo in questo paragrafo, al punto tale che qualche hanno fa, nell'ambito di una importante manifestazione scientifica, il lievito è stato nominato mammifero onorario!

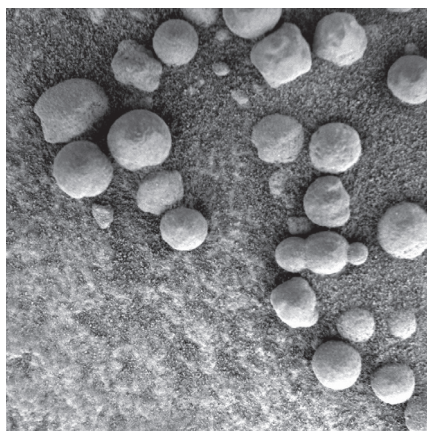
Tuttavia, altre specie di lievito sono utilizzate in laboratorio per studi di Genetica e Biologia Molecolare. Tra queste, il lievito *Schizosaccharomyces pombe*. A differenza di *S. cerevisiae*, che si divide per gemmazione (*budding*), *S. pombe* si divide per scissione mediana (**fig. 22.1**). Il primo ha una forma ellissoidale, il secondo ha una forma a bastoncino.

Il lievito è un eucariote unicellulare e ciò rappresenta la caratteristica più rilevante del suo successo come organismo modello, offrendo numerosi vantaggi come sistema sperimentale. I lieviti hanno genomi piccoli rispetto ad altri eucarioti e, quindi, un altrettanto piccolo numero di geni. *S. cerevisiae* è il primo organismo eucariotico di cui sia stata determinata la sequenza nucleotidica dell'intero genoma e contiene circa 6000 geni. La dimensione del suo genoma è di 12 Mpb, circa 3 volte più grande del batterio *E. coli*, ma circa 260 volte più piccolo del genoma dell'uomo. Proprio per questo motivo, i 16 cromosomi di *S. cerevisiae* sono di dimensioni ridotte e poco adatti a un'analisi citogenetica al microscopio; sono, però, analizzabili con tecniche molecolari più analitiche. Infatti, gli interi cromosomi di lievito possono essere separati e analizzati mediante particolari tipi di elettroforesi su gel.

I lieviti vengono coltivati in laboratorio in modo molto semplice, analogamente a quanto si può fare con i batteri. Nonostante la loro semplicità, le cellule di *S. cerevisiae* hanno le stesse caratteristiche generali di eucarioti pluricellulari. I lieviti, infatti, hanno un nucleo distinto circondato da un involucro nucleare e

Figura 22.1 Fotografie al microscopio elettronico del lievito gemmante *Saccharomyces cerevisiae* (A) e del lievito *Schizosaccharomyces pombe* (B).

A



B



il loro citoplasma è dotato dell'intero corredo di organelli cellulari (mitocondri, vacuoli, apparato di Golgi ecc.) presenti in altre cellule eucariotiche. Il lievito *S. cerevisiae*, oltre che da una tipica membrana cellulare, è circondato all'esterno da una parete cellulare piuttosto resistente.

Da migliaia di anni, il lievito *S. cerevisiae* è stato "addomesticato" dall'uomo per la produzione di pane, vino e birra. È quindi un organismo utilizzato per una serie di "processi biotecnologici naturali" di cui oggi si conoscono i meccanismi chimici e molecolari.

Come sopra accennato, le cellule di *S. cerevisiae* sono di forma ellissoidale con un diametro di circa 5 micron e si dividono per gemmazione, una caratteristica che ha reso tale lievito molto utile per studi sui meccanismi di divisione cellulare.

S. cerevisiae può crescere sia allo stato aploide (cioè con una copia di ciascun cromosoma e quindi, in genere, con una copia di ciascun gene) che allo stato diploide (cioè con due copie di ciascun cromosoma). La conversione tra lo stato aploide a quello diploide è mediata dall'accoppiamento tra due ceppi di lievito e quella dallo stato diploide ad aploide da un processo chiamato **sporificazione** (fig. 22.2).

I ceppi di lievito aploidi esistono in due tipi sessuali distinti detti a e α . Ceppi di lievito, chiamati **eterotallici**, di tipo sessuale a o α possono essere fatti crescere come ceppi aploidi stabili per indefinite generazioni. Solo se i due tipi sessuali opposti a e α vengono fatti crescere insieme, le cellule di sesso opposto si accoppiano generando una cellula diploide a/α che, in certe condizioni, può dividersi indefinitamente mantenendo lo stato diploide.

Ceppi di lievito diploidi possono essere fatti sporificare se vengono messi in condizioni limitanti di nutrienti necessari per la crescita. In tali condizioni, le cellule che stavano dividendosi mitoticamente iniziano invece il processo della meiosi, alla fine del quale vengono prodotte quattro cellule aploidi, due di sesso a e due di sesso α , che sono i gameti chiamati anche, in questo caso, **spore**. Le quattro spore sono contenute all'interno di una struttura chiamata **asco** (visibile

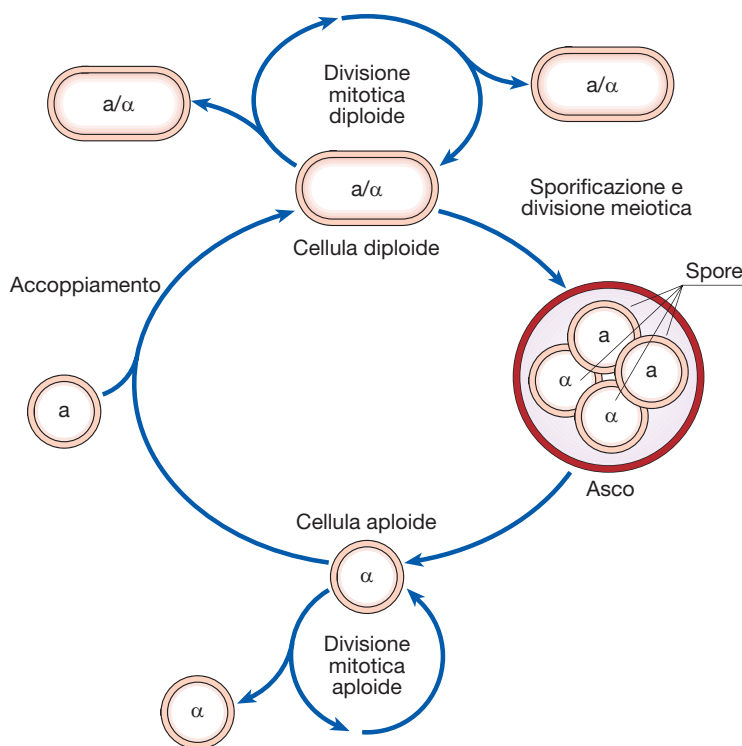
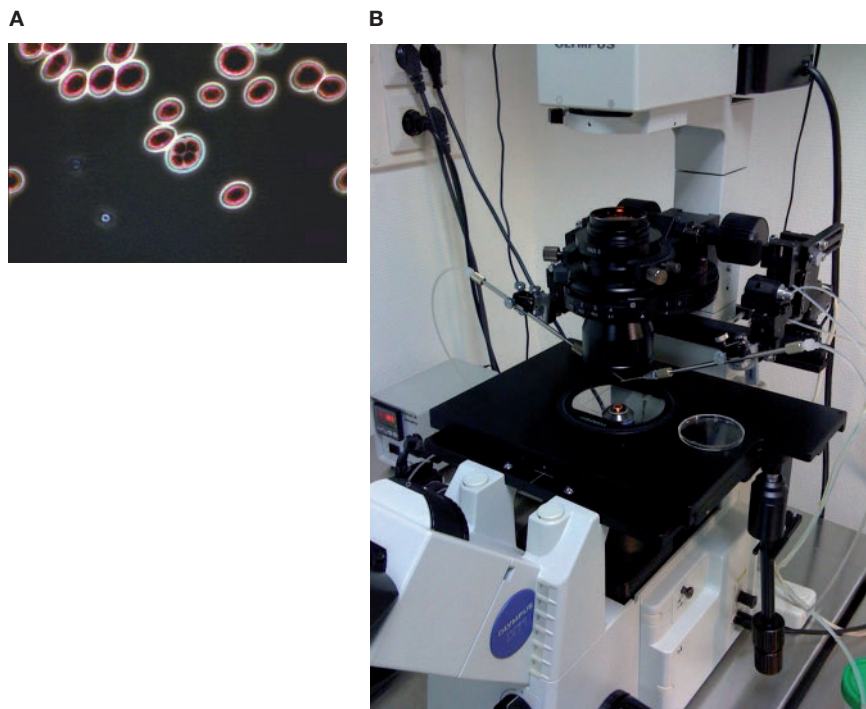


Figura 22.2 Ciclo vitale del lievito *Saccharomyces cerevisiae*. I due ceppi aploidi (a e α) possono accoppiarsi e formare il ceppo diploide a/α che, in condizioni limitanti di nutrienti, va incontro a sporificazione rigenerando nel processo meiotico spore aploidi.

Figura 22.3 Dissezione di un asco di *Saccharomyces cerevisiae*. (A) Nel centro della figura è visibile un asco in cui si distinguono le quattro spore aploidi prodotte durante il processo meiotico. (B) Un microscopio con micromanipolatore utilizzato per la dissezione degli aschi.



al microscopio), che racchiude quindi i gameti prodotti da una singola meiosi (**fig. 22.3A**).

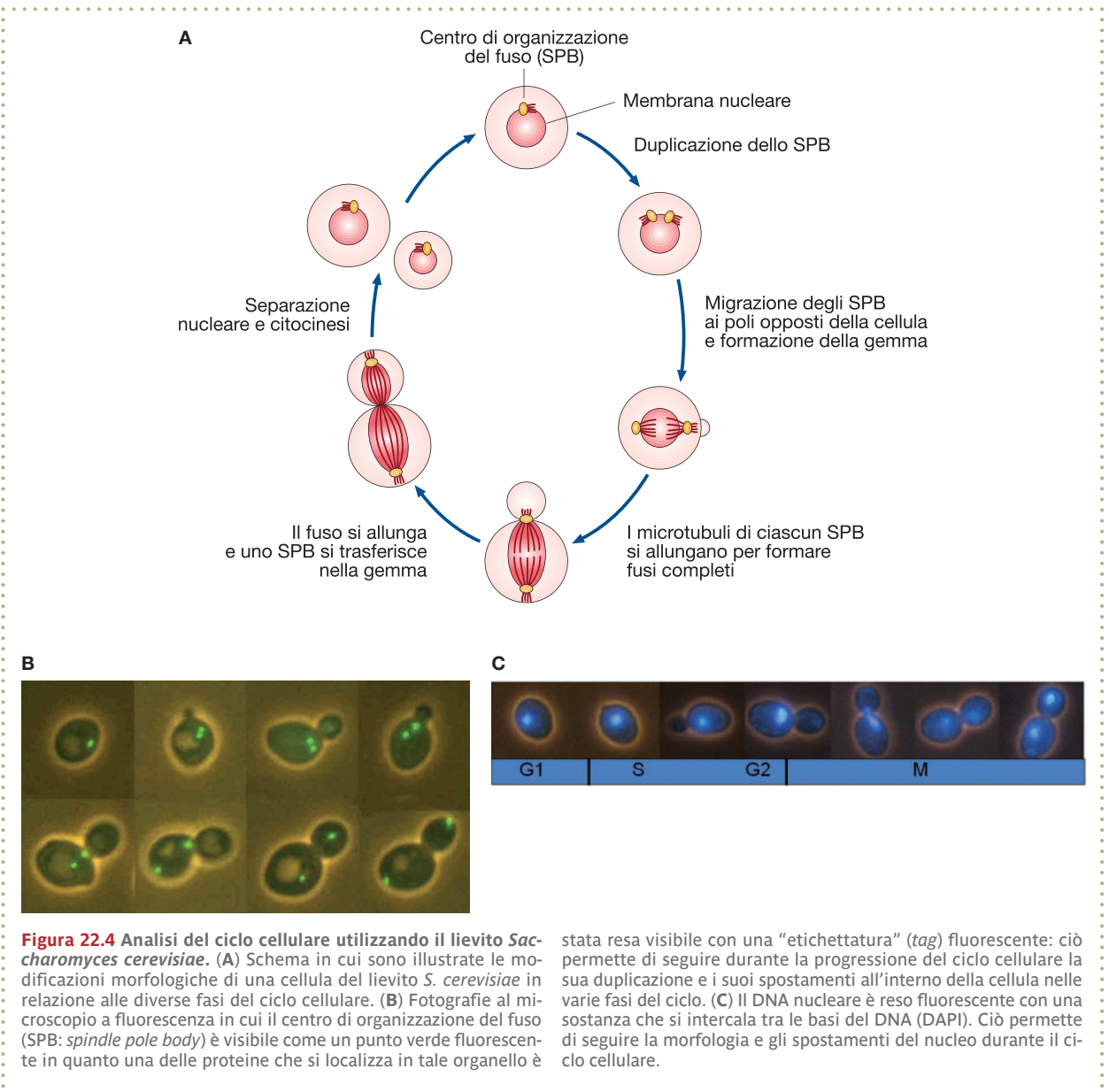
Il lievito ha quindi la peculiarità che i gameti prodotti in un singolo evento meiotico sono contenuti in una specifica struttura e sono analizzabili singolarmente. Ciò è impossibile per quasi tutti gli altri organismi eucariotici e ha reso il lievito l'organismo d'elezione per l'analisi genetica della segregazione dei caratteri e per studi sui meccanismi di ricombinazione. I gameti prodotti durante ogni evento meiotico possono essere prelevati con un apparecchio, chiamato micromanipolatore (**fig. 22.3B**), che è un microscopio con attaccato un ago che possiamo muovere in modo controllato così da andare a rompere gli aschi e prelevare le singole spore (gameti) in esso contenute.

Le spore vengono poi posizionate su terreni di coltura solidi e ciascun gamete aploide si dividerà generando una colonia di cellule. Tutte le cellule che formano la colonia hanno lo stesso genotipo ed è così possibile, utilizzando opportuni terreni e condizioni di crescita, valutare come un certo carattere segrega durante la meiosi.

L'utilizzo del lievito *S. cerevisiae* in studi di Biologia Molecolare è stato rilevante per comprendere molteplici processi biologici. Fra questi, però, vogliamo qui ricordare la comprensione dei meccanismi molecolari del controllo del ciclo cellulare (*vedi* anche par. 6.7). Per tali studi Leland Hartwell e Paul Nurse, insieme con Timothy Hunt che lavora su *Xenopus laevis*, hanno ottenuto nel 2003 il premio Nobel per la fisiologia e la medicina.

Perché lo studio del ciclo cellulare del lievito *S. cerevisiae* è stato così importante?

Il primo motivo è legato alla modalità con cui tale lievito si divide. Infatti, poiché *S. cerevisiae* si divide per gemmazione, è possibile distinguere facilmente in che fase del ciclo cellulare una cellula si trovi, semplicemente guardando la morfologia della cellula stessa al microscopio (**fig. 22.4A**). Come indicato nello schema in figura, cellule di lievito in fase G1 hanno una forma ellissoidale; quando entrano nella fase S cominciano a emettere la gemma, che diventa via



via sempre più grossa. Quando le cellule entrano nella fase G2, il nucleo singolo presente nella cellula madre si posiziona tra madre e figlia, si allunga poi tra le due cellule e viene segregato e separato in mitosi. Se si colora il DNA nucleare con un opportuno colorante (**fig. 22.4C**), la posizione del nucleo nelle diverse fasi del ciclo cellulare risulta molto chiara. Analogamente, se si evidenzia con opportuni trucchi una proteina che si localizza specificatamente nel centro dell’organizzazione del fuso (equivalente al cinetocoro nelle cellule di mammifero), la dinamica della duplicazione e localizzazione nel ciclo cellulare di tale organello diventa evidente (**fig. 22.4B**).

Un aspetto a cui vogliamo accennare per spiegare l’utilità del lievito per la comprensione del ciclo cellulare riguarda l’essenza delle ricerche che hanno fruttato a Hartwell e Nurse l’assegnazione del premio Nobel. Poiché l’incapacità di progredire attraverso il ciclo cellulare è letale per la cellula, Hartwell e Nurse cercarono mutazioni che bloccassero la progressione del ciclo cellulare di lievito solo in particolari condizioni: per esempio, mutazioni temperatura-sensibili (*vedi*

cap. 6). Tali mutazioni hanno permesso l'identificazione dei geni che controllano la progressione del ciclo cellulare, che vennero chiamati **geni CDC** (*Cell Division Cycle*). Tali geni sono stati poi identificati utilizzando le tecnologie del DNA ricombinante descritte nel capitolo 20 e, in particolare, la tecnica della "complementazione funzionale".

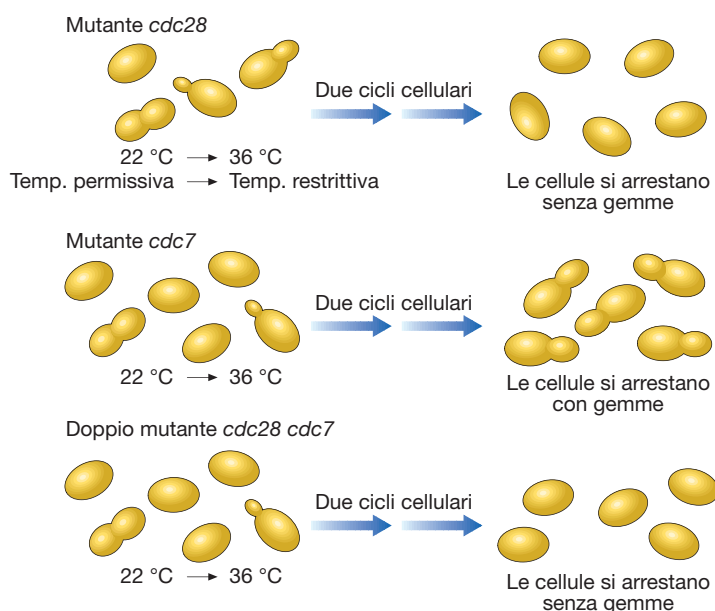
Come mostrato in **figura 22.5**, l'individuazione di mutanti *cdc* in *S. cerevisiae* è facilitata dal fatto che questo lievito possiede una serie di parametri morfologici che possono essere facilmente seguiti durante il ciclo cellulare. Mutazioni temperatura-sensibili nei geni *CDC* che controllano uno specifico passaggio del ciclo cellulare sono identificabili perché le cellule mutanti si arrestano tutte con la stessa morfologia (**fenotipo terminale**), dopo il trasferimento alle condizioni non-permissive, in quanto non sono capaci di eseguire una funzione richiesta per il passaggio da uno stadio all'altro del ciclo.

Combinando poi nello stesso ceppo due mutazioni *cdc* e osservando il fenotipo terminale si riesce a stabilire l'ordine temporale d'azione dei diversi geni *CDC* (**fig. 22.5**). Ciò ha permesso di posizionare tutti i geni *CDC* in una sequenza temporale durante il ciclo cellulare.

Gli studi successivi hanno dimostrato che i geni *CDC*, scoperti in lievito, sono conservati in tutti gli eucarioti e hanno permesso di comprendere la maggior parte dei meccanismi molecolari che controllano la divisione e proliferazione cellulare.

L'ultimo aspetto che vale la pena di ricordare è che il lievito è l'organismo che ha aperto la strada agli studi di Genomica funzionale e della Genetica inversa a cui abbiamo già accennato nel capitolo 20 (*vedi* finestra 20.5 e par. 20.10). Con il termine **Genomica funzionale** si intende lo studio della funzione di tutti i geni di un organismo e con il termine **Genetica inversa** si intende un approccio sperimentale che è opposto a quello utilizzato da Mendel. Come tutti sanno, Mendel studiò come alcuni caratteri osservabili in piante di pisello segregano di generazioni in generazione. Oggi sappiamo che i caratteri analizzati da Mendel sono controllati da specifici geni e che mutazioni in tali geni sono alla base dei fenotipi la cui segregazione fu analizzata da Mendel. Attraverso il sequenziamento d'interi

Figura 22.5 Utilizzo di mutanti *cdc* di *S. cerevisiae* per determinare la successione di diversi eventi molecolari durante il ciclo cellulare. Il gene *CDC28* codifica per una funzione richiesta per il passaggio dalla fase G1 alla fase S. Infatti, tutte le cellule *cdc28* mutanti si bloccano con la morfologia tipica di cellule in G1 (senza gemma) dopo trasferimento alla temperatura restrittiva. Il gene *CDC7* codifica una funzione richiesta per l'inizio della fase S. Infatti, tutte le cellule mutanti *cdc7* hanno la stessa morfologia (una gemma più piccola della cellula madre, tipica di cellule in fase S). Cellule che portano entrambe le mutazioni si bloccano con la morfologia tipica dei mutanti *cdc28* e non di mutanti *cdc7*, indicando che il gene *CDC28* agisce temporalmente prima del gene *CDC7*.



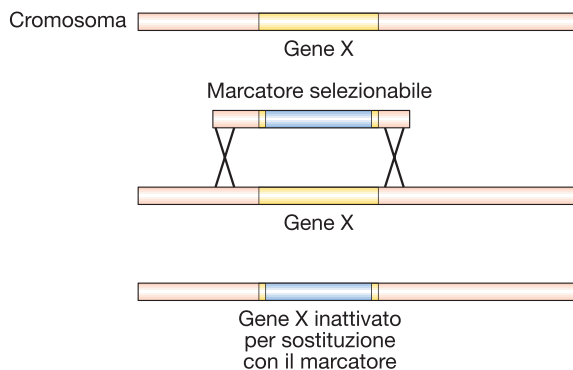
Conclusione: *CDC28* funziona prima di *CDC7*

genomi conosciamo il genotipo completo di un organismo. La Genetica inversa, attraverso l'eliminazione della funzione dei diversi geni, o la produzione di specifici mutanti, si prefigge di capire la funzione di ciascun gene di un organismo.

Abbiamo già descritto nel capitolo 8 i meccanismi della ricombinazione genetica. Quando un frammento di DNA esogeno viene introdotto all'interno di una cellula tramite trasformazione, se non può replicarsi, viene generalmente degradato dalla cellula stessa. In una piccola percentuale di casi, però, tale frammento di DNA può integrarsi tramite fenomeni di ricombinazione all'interno dei cromosomi della cellula ospite. Nella maggior parte dei casi (ad es. in cellule di mammifero) tale integrazione avviene a caso, secondo meccanismi detti di **ricombinazione illegittima**, che non richiedono omologie di sequenza nucleotidica. Nel lievito, la maggior parte degli eventi di integrazione avviene, invece, attraverso dei meccanismi di **ricombinazione omologa** tra sequenze presenti sul tratto di DNA inserito nella cellula tramite trasformazione e sequenze omologhe sui cromosomi di lievito.

Come mostrato in **figura 22.6**, per inattivare un gene X di nostro interesse si introduce in una cellula diploide di lievito un frammento di DNA che contenga, ai lati di una marcatore selezionabile (come il gene *URA3* discusso nel capitolo 20, qui indicato dal colore azzurro) sequenze identiche alle zone fiancheggianti il gene che si vuole “distruggere”. Il frammento introdotto nella cellula tramite trasformazione non può replicarsi perché privo di un'origine di replicazione del DNA e, quindi, il suo destino è quello di venire perso (nella maggior parte dei casi) o di inserirsi all'interno del cromosoma di lievito mediante un doppio crossing over. Questo è un evento raro, ma selezionabile dal fatto che le poche

A



B

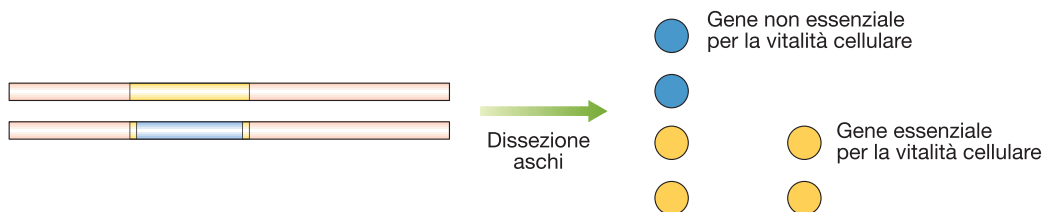


Figura 22.6 Esperimento di “distruzione genica” in lievito. (A) Un gene X (in giallo) su uno dei 16 cromosomi di *S. cerevisiae* può venire inattivato mediante l'introduzione all'interno del nucleo di cellule di lievito di un frammento di DNA in cui parte del gene X (in giallo) è stata sostituita dal marcatore selettivo (in azzurro), la cui presenza conferisce alle cellule che lo contengono un fenotipo facilmente individuabile (vedi testo). Questo frammento di DNA può sostituire il gene X sul cromosoma determinando l'inattivazione della sua funzione (distruzione genica) e l'espressione del marcatore selettivo.

(B) L'esperimento descritto nel pannello (A) viene eseguito su un ceppo diploide di lievito in modo da provocare la distruzione di una soltanto delle due copie del gene X. Dopo sporificazione e dissezione degli aschi per analizzare il genotipo delle spore, si ottengono quattro spore vive se il gene X inattivato non svolge una funzione essenziale per la vitalità cellulare e due di queste (in azzurro) presenteranno il fenotipo associato alla presenza del marcatore. Se il gene X codifica per una funzione essenziale, solo le due spore di ogni asco che non hanno integrato il marcatore saranno vitali.

cellule che sono andate incontro a tale processo sono diventate Ura^+ , perché hanno introdotto il gene marcatore *URA3*. Tali cellule saranno positivamente selezionabili su un terreno privo di uracile, se l'esperimento è stato eseguito con mutanti auxotrofi Ura^- . L'evento d'integrazione porta all'inattivazione del gene, dato che il gene X è stato interrotto dall'integrazione del gene marcatore *URA3*. Se l'esperimento è stato eseguito in modo tale che l'integrazione avvenga su una sola delle 2 copie del gene X, riusciamo anche a valutare se il gene X codifica o meno per una funzione necessaria per la vita della cellula. Se, infatti, facciamo sporificare il ceppo diploide che contiene la distruzione di una copia del gene X e il gene codifica per una funzione essenziale, ci aspettiamo di trovare solo 2 spore vive in grado di dare origine a una colonia. Viceversa, se il gene X codifica per una funzione non essenziale per la vitalità cellulare, tutte e 4 le spore daranno origine a colonie (**fig. 22.6**). Questa procedura è stata applicata sistematicamente a tutti i circa 6000 geni di lievito e si è cominciato a capire quanti sono i geni essenziali e quanti quelli non essenziali.

Esperimenti analoghi a quello sopra descritto possono oggi essere compiuti anche in mammiferi, ma la loro esecuzione è estremamente più difficile che in lievito, perché nei mammiferi la maggior parte degli eventi di integrazione non avvengono su sequenze omologhe di DNA. Per la costruzione di animali transgenici nei mammiferi è, perciò, stato necessario sviluppare nuove strategie che saranno brevemente discusse fra poco.

22.2 Il moscerino della frutta *Drosophila melanogaster*

Sono passati oltre cento anni da quando nel 1908 Thomas H. Morgan (**fig. 22.7**) e il suo gruppo alla Columbia University selezionarono *Drosophila melanogaster* come un organismo modello. Avevano messo della frutta quasi marcia sul davanzale della finestra del laboratorio cercando dei piccoli animali che si riproducevano in fretta e che potessero essere usati in laboratorio per l'analisi dell'ereditarietà di caratteri quantitativi. È oggi sorprendente sapere, attraverso studi di biogeografia storica, che *Drosophila* si era diffusa negli Stati Uniti solo in tempi recenti. *Drosophila* è una specie che ha la sua origine nell'Africa equatoriale, ma si è diffusa nel resto del mondo seguendo la colonizzazione del pianeta da parte dell'uomo. *Drosophila* è, infatti, molto vulnerabile in un ambiente naturale, mentre deve il suo successo come "commensale" delle attività dell'uomo: infatti, convive facilmente in insediamenti umani e utilizza diverse fonti di cibo riproducendosi in modo rapido. Sotto questo aspetto assomiglia moltissimo a un altro organismo modello, il topo *Mus musculus*, che descriveremo più avanti e che si è anch'esso diffuso su tutto il pianeta come "commensale" delle attività umane.

Praticamente non c'è aspetto della ricerca biologica moderna e anche della Genetica classica in cui non si sia fatto uso di *Drosophila melanogaster* come modello. Ci vorrebbero interi testi di Genetica e di Biologia Molecolare per rendere omaggio a questo piccolo moscerino per tutte le conoscenze che ci ha permesso di acquisire negli ultimi cento anni. Qui ci limiteremo, però, a riassumere le caratteristiche salienti del ciclo vitale di *Drosophila* e a citare le aree di ricerca in cui il suo utilizzo è stato più rilevante.

I moscerini della specie *Drosophila melanogaster* hanno occhi rossi, un corpo giallo-marrone con l'addome attraversato da alcuni anelli scuri (**fig. 22.8**). Il primo mutante trovato dal gruppo di Morgan aveva un fenotipo evidente: aveva gli occhi bianchi. Nella specie c'è un chiaro dimorfismo sessuale: le femmine sono più grandi (circa 2,5 mm), mentre i maschi, più piccoli, hanno la parte terminale più scura e un ciuffo di peli che circonda la chiusura dell'addome e gli organi sessuali.

A 25 °C il ciclo vitale di *Drosophila* è di circa due settimane (**fig. 22.9**). Le femmine riescono a deporre fino a 600 uova e le uova si schiudono dopo un



Figura 22.7 Thomas H. Morgan (1888-1945). Morgan ha introdotto, intorno al 1910, l'uso di *Drosophila melanogaster* come organismo modello per lo studio della Genetica.



Figura 22.8 Il moscerino della frutta *Drosophila melanogaster*. Fotografie di un maschio (destra) e di una femmina (sinistra) di *Drosophila melanogaster*.

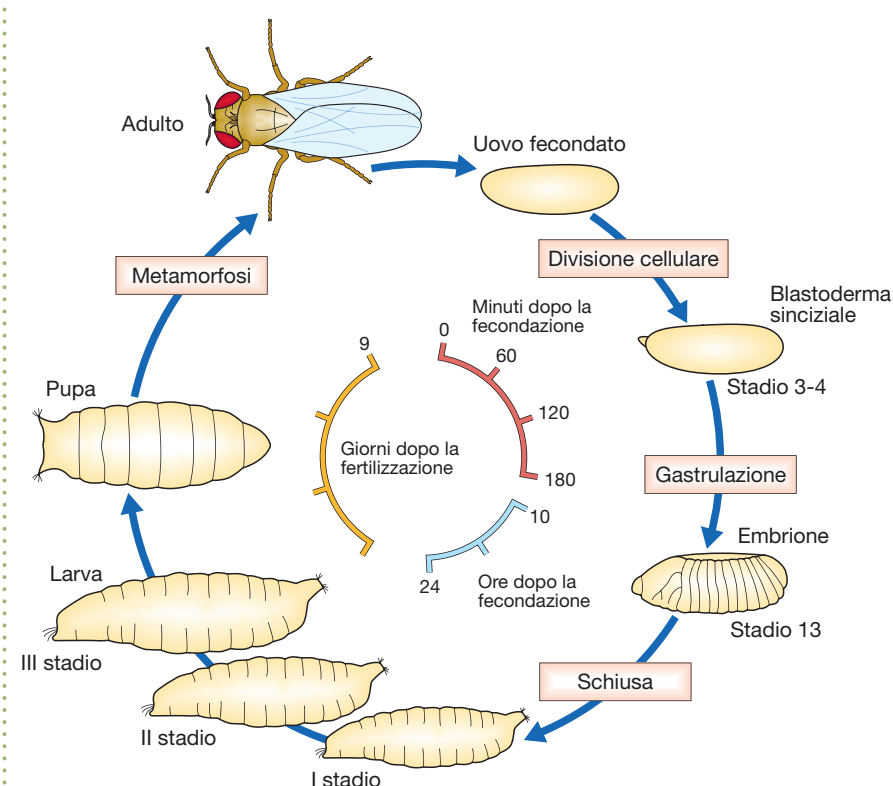
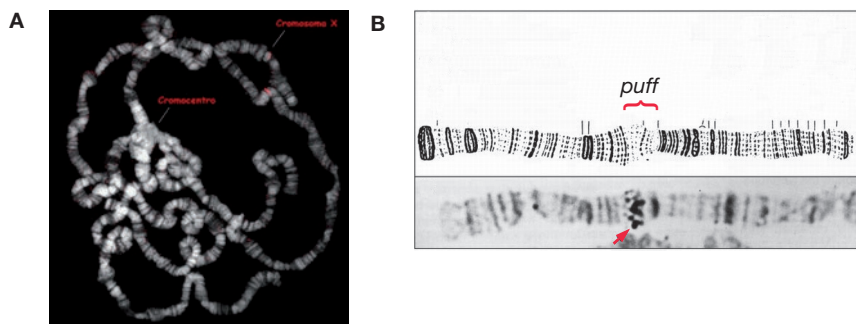


Figura 22.9 Ciclo vitale di *Drosophila melanogaster*.

giorno dalla deposizione originando delle larve al primo stadio. Le larve crescono per alcuni giorni, passando da uno stadio all'altro e trasformandosi, infine, in una "pupa": da quest'ultimo stadio, attraverso un processo chiamato **metamorfosi**, in circa cinque giorni si sviluppa l'insetto adulto. Le femmine sono in grado di accoppiarsi poche ore dopo la metamorfosi accumulando lo sperma in specifiche sacche che verranno successivamente utilizzate per fecondare le uova. Per questo motivo i genetisti separano le femmine dal resto della popolazione prima che possano accoppiarsi, così che l'incrocio avvenga solo con il maschio con il genotipo e fenotipo prescelto per l'esperimento. Gli insetti sono facilmente cresciuti in laboratorio in bottiglie contenenti come cibo una specie di pappa costituita, per lo più, da lievito e agar.

Drosophila melanogaster ha quattro paia di cromosomi: una coppia X/Y e tre coppie di autosomi denominate 2, 3 e 4. Come abbiamo già visto (*vedi* parr. 4.5

Figura 22.10 I cromosomi politenici delle ghiandole salivari di *Drosophila*. (A) Immagine al microscopio dei cromosomi. (B) Sopra: rappresentazione schematica di un tratto di cromosoma politenico in cui è indicato il *puff* corrispondente al gene per HSP70. Sotto: lo stesso tratto di cromosoma su cui è stata eseguita una ibridazione *in situ* con una sonda per l'mRNA di HSP70.



e 6.8), nelle ghiandole salivari delle larve i cromosomi vanno incontro a dei cicli di re-duplicazione del DNA successivi non seguiti da separazione dei cromosomi figli. Si generano, così, dei **cromosomi** giganti, chiamati **politenici**, in cui è possibile distinguere bande e interbande regolari che permettono un'analisi genetica accurata e, perfino, un'analisi citologica dell'espressione genica, dato che durante l'embriogenesi regioni trascritte dei cromosomi politenici sono visibili come dei rigonfiamenti detti *puff* (fig. 22.10).

Senza entrare in dettagli, vogliamo qui presentare una breve retrospettiva storica dei maggiori contributi che il moscerino della frutta ha dato allo sviluppo della Genetica e Biologia Molecolare moderna.

Nel 1910 Morgan isolò il primo mutante “occhi bianchi” di *Drosophila* e il suo studente Alfred H. Sturtevant nel 1913 costruì la prima mappa genetica che dimostrò che i geni sono allineati linearmente sui cromosomi. Calvin B. Bridges, un altro allievo di Morgan, utilizzando il meccanismo di non-disgiunzione cromosomica in femmine XXY, offrì la prima elegante prova che i cromosomi contengono i geni, un concetto che, a quell'epoca, era molto controverso. Tutti questi studi portarono all'assegnazione del premio Nobel a Morgan nel 1933 per la “teoria cromosomica dei meccanismi dell'ereditarietà”. Ancora un altro studente di Morgan, Hermann J. Muller, dimostrò come le radiazioni ionizzanti sono in grado d'indurre danni al DNA e provocare mutazioni e aberrazioni cromosomiche, aprendo la strada agli studi di mutagenesi e permettendogli di ottenere il premio Nobel nel 1946. Negli anni dal 1935 al 1938 Bridges pubblicò il mappaggio fisico dei geni sui cromosomi politenici con un'accuratezza utilizzata ancora ai nostri giorni.

Nel 1974 David S. Hogness costruì la prima libreria di DNA utilizzando DNA di *Drosophila* e sviluppò i primi metodi di screening per individuare cloni di DNA corrispondenti a specifiche sequenze tramite la tecnica di “ibridazione su colonia” (vedi cap. 20), mentre nel 1978 Tom Maniatis inventò la prima tecnica che permette di spostarsi lungo un cromosoma (*chromosome walking*) utilizzando segmenti di DNA cromosomico parzialmente sovrapposti in librerie costruite nel batteriofago lambda.

Nel 1980 Christiane Nusslein-Volhard ed Eric Wieschaus riuscirono per la prima volta a trasferire in un organismo multicellulare l'analisi mutazionale a livello di tutto il genoma per individuare i geni richiesti per un processo fondamentale degli eucarioti quale lo sviluppo embrionale. I due ricercatori ottennero il premio Nobel nel 1995. Nel 1982 Allan C. Spradling e Gerald M. Rubin svilupparono la prima metodologia per ottenere dei moscerini transgenici tramite l'utilizzo di elementi trasponibili. Gli sviluppi recenti di questo approccio sperimentale hanno portato negli ultimi anni all'acquisizione di tecnologie sempre più sofisticate quali:

- l'utilizzo di trasposoni, in particolare degli elementi P, per generare una serie di distruzioni geniche e analizzarne il fenotipo;

- la possibilità di eseguire esperimenti di ricombinazione sito-specifica in condizioni controllabili;
- la possibilità di identificare rapidamente ed efficacemente cambiamenti di singoli nucleotidi in specifici geni (*tilling*);
- la possibilità di poter sovra-esprimere geni con il sistema Gal4 sviluppato in lievito ecc.

Le potenzialità di queste tecnologie sono state esaltate dal sequenziamento del genoma di *Drosophila melanogaster*, ottenuto nel 2000, che ha portato al riconoscimento di circa 14 000 geni.

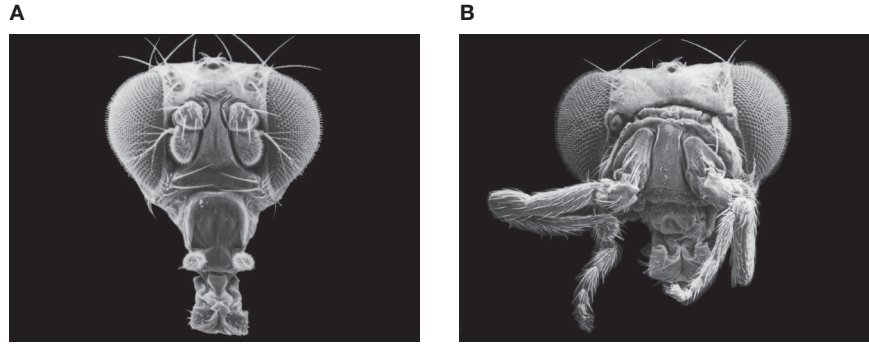
In quali ambiti della ricerca *Drosophila melanogaster* ha permesso di ampliare le nostre conoscenze? Già dall'elenco delle più importanti scoperte che tale organismo ha permesso di ottenere, la risposta a questa domanda non può che essere che *Drosophila* è stato un modello sperimentale per gli eucarioti multicellulari essenzialmente in tutti gli ambiti della ricerca. Se proprio si vuole delimitare un singolo ambito più specifico suggeriremmo l'area della Genetica dello sviluppo, che, per quanto riguarda *Drosophila*, cerchiamo di riassumere in poche righe.

La vita nel moscerino inizia come nell'uomo: uno spermatozoo feconda una cellula uovo, formando un unico nucleo diploide. Questo nucleo, in *Drosophila*, inizia a dividersi rapidamente formando un sincizio, cioè una singola cellula multinucleata. Spostandosi via via dalla regione centrale verso la corteccia dell'uovo, i nuclei formano un monostrato di circa 6000 nuclei che circondano il tuorlo centrale. Nel tempo di 1-3 ore dalla fecondazione si formano le membrane cellulari tra i nuclei adiacenti. Mentre prima della formazione della membrana i nuclei sono totipotenti, cioè capaci di dare origine a qualunque tipo di cellula, dopo la formazione della membrana diventano "determinati" a differenziarsi nei tessuti specifici del moscerino. Gli screening genetici di Nusslein-Volhard e Wieschaus portarono all'identificazione dei geni che controllano la suddivisione precoce dell'embrione di *Drosophila*. Questi geni controllano la formazione di gradienti di morfogeni che portano alla determinazione degli assi antero-posteriore e dorso-ventrale dell'embrione. Si scoprì poi che la maggior parte di questi geni sono fattori trascrizionali le cui concentrazioni possono attivare o reprimere particolari geni nello zigote.

Il destino delle diverse cellule in *Drosophila* si realizza in due tempi: la **specificazione** e la **determinazione**. All'inizio dello sviluppo il destino della cellula dipende da fattori ambientali, quale l'instaurarsi di gradienti di proteine come descritto sopra. Questa "specificazione" delle cellule è però ancora flessibile e può essere modificata da segnali che provengono da altre cellule. Successivamente nello sviluppo, questa "specificazione" diventa "determinata" e irreversibile. Tale transizione è geneticamente controllata dai **geni della segmentazione**: questi geni, per esempio, dividono l'embrione primitivo in una serie di segmenti specifici lungo l'asse antero-posteriore. I *gap genes*, ancora una volta, vennero scoperti negli studi di Nusslein-Volhard e Wieschaus.

Successivamente alla definizione dei confini dei diversi segmenti del corpo dell'embrione, devono venire stabilite le strutture caratteristiche di ciascun segmento. Questa specificazione avviene a opera dei **geni selettori omeotici**. Poiché questi ultimi sono responsabili delle strutture presenti nelle diverse parti del corpo del moscerino, mutazioni in tali geni portano a fenotipi bizzarri identificati con nomi strani, come accade per tutte le mutazioni di *Drosophila*. Per esempio, il corpo del moscerino adulto normale comprende tre segmenti toracici, ciascuno dei quali forma un paio di zampe. Il primo segmento toracico non forma altre appendici, mentre il secondo segmento toracico, oltre alle zampe, forma un paio di ali. Il terzo segmento toracico forma un paio di zampe e un paio di bilancieri. In mutanti omeotici queste specifiche identità dei segmenti possono essere modificate. In caso di delezione del gene *ultrabithorax*, il terzo segmento toracico

Figura 22.11 I mutanti di *Drosophila* nello studio dello sviluppo. Differenze morfologiche tra la testa di una *Drosophila* selvatica (A) e del mutante di *Drosophila* chiamato *antennapedia* (B).



(caratterizzato dai bilancieri) è trasformato in un secondo segmento toracico; il risultato è un moscerino con quattro ali. Analogamente, la proteina Antennapedia specifica di solito il secondo segmento toracico del moscerino. Ma quando il moscerino reca una mutazione in cui il gene *antennapedia* viene espresso nel capo (oltre che nel torace), dal capo crescono zampe invece che antenne (**fig. 22.11**). I geni omeotici agiscono attivando o reprimendo un gruppo di “geni realizzatori”, che sono i bersagli delle proteine dei geni omeotici e che funzionano formando specifici tessuti o abbozzi di organi.

Oltre che per studi di genetica dello sviluppo, *Drosophila* si è rivelato un organismo modello estremamente utile per lo studio di patologie nell'uomo. Sulla base di dati abbastanza recenti di circa 2300 geni umani probabilmente connessi a patologie nell'uomo, oltre 700 hanno chiari omologhi funzionali in *Drosophila*. Analogamente, *Drosophila* ha geni omologhi all'uomo che, se inattivati, causano un ampio spettro di patologie neurologiche, tumori, alterazioni nello sviluppo, malattie metaboliche e vascolari, così come malattie collegate ad alterazioni della vista, dell'udito e del sistema immunitario. Ne consegue che *Drosophila* è spesso utilizzata come un sistema modello complesso multicellulare per analizzare in modo sistematico la funzione di un numero elevato di geni possibilmente coinvolti nel causare una serie di patologie nell'uomo.

22.3 Il nematode *Caenorhabditis elegans*



Figura 22.12 Sidney Brenner (1927-). Brenner ha introdotto, nel 1963, il nematode *Caenorhabditis elegans* come organismo modello per lo studio della Biologia dello sviluppo e della Neurobiologia dei vertebrati.

Sidney Brenner (**fig. 22.12**) per primo propose circa quarant'anni fa *Caenorhabditis elegans* come modello sperimentale per studi nell'ambito della Neurobiologia e della Biologia dello sviluppo. Oggi *C. elegans* è utilizzato per affrontare un numero molto più ampio di problemi biologici di grande interesse, quali la trasduzione del segnale, il controllo del ciclo cellulare e l'apoptosi, la polarità cellulare, il processo d'invecchiamento e la determinazione del sesso. Moltissime scoperte basilari, sia nell'ambito della ricerca di base che in aree importanti per la ricerca biomedica, sono state ottenute utilizzando questo piccolo nematode. Basti ricordare che nel 2002 Sidney Brenner, Robert Horvitz e John Sulston hanno vinto il premio Nobel per la fisiologia e la medicina “per le loro scoperte sulla regolazione genetica dello sviluppo degli organi e della morte cellulare programmata” e lo stesso premio è stato assegnato nel 2006 ad Andrew Fire e Craig Mello “per la loro scoperta del meccanismo dell'interferenza dell'RNA nel silenziamento genico”.

C. elegans ha diverse caratteristiche che lo rendono un ottimo organismo modello. Per prima cosa, è estremamente facile da tenere in laboratorio: in natura, il piccolo verme vive nel suolo e si ciba di vari batteri, ma può essere cresciuto in laboratorio a 20-25 °C su capsule Petri nutrendolo con cellule di *Escherichia coli*. Inoltre, si riproduce rapidamente ed è molto prolifico: in 3 giorni dall'uovo



Figura 22.13 Il piccolo nematode *Caenorhabditis elegans*.

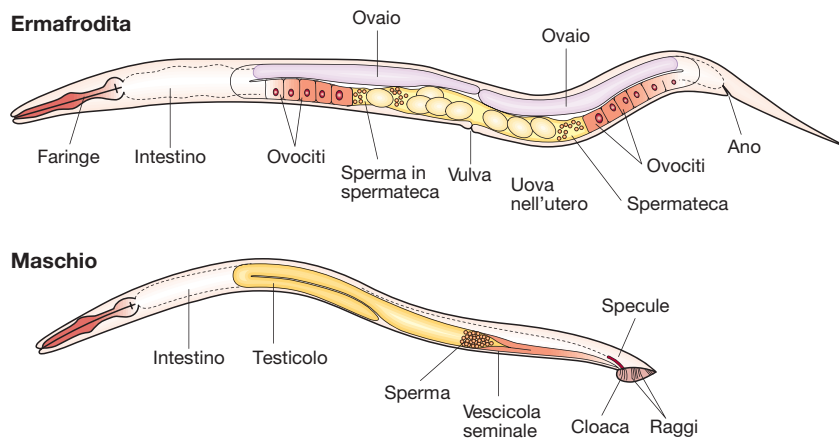


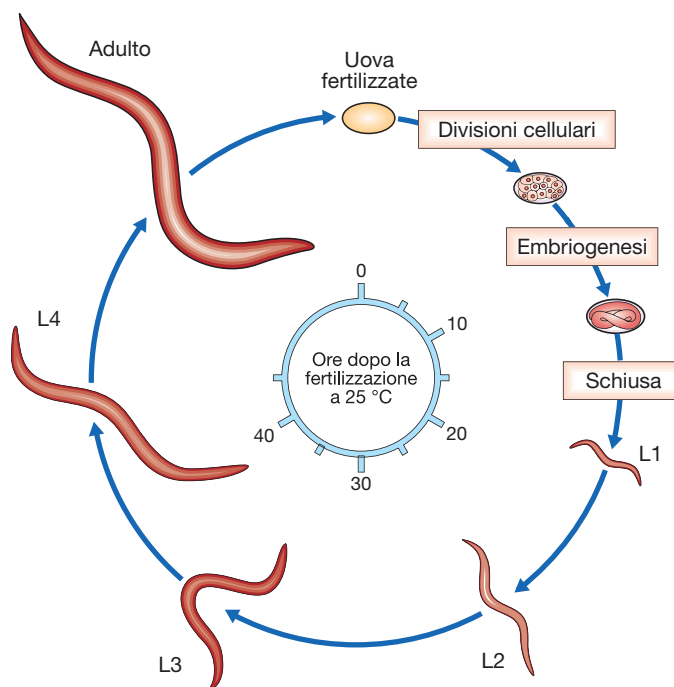
Figura 22.14 Anatomia del maschio e dell'ermafrodita di *Caenorhabditis elegans*.

si forma un organismo adulto delle dimensioni di circa 1,3 mm in lunghezza (**fig. 22.13**).

In una popolazione di *C. elegans* sono presenti organismi di due tipi sessuali diversi: il maschio e l'ermafrodita (**fig. 22.14**). I maschi sono molto rari (pari a circa 0,05% del totale) e garantiscono la variabilità genetica. Gli ermafroditi sono in grado di proliferare anche senza maschi e sono, quindi, molto numerosi. Oltre all'apparato maschile, essi presentano anche due ovari, due ovidotti, una spermateca e un utero. Il corredo cromosomico di *C. elegans* comprende cinque paia di autosomi e due cromosomi sessuali: l'organismo ermafrodita ha due cromosomi X, mentre i maschi presentano una combinazione XO. L'aspetto fisico dell'ermafrodita è differente rispetto a quello del maschio, dal momento che è dotato di una coda più lunga; il maschio presenta invece una coda tagliata e un'area laterale a forma di ventaglio.

Le uova di *C. elegans* sono deposte dall'ermafrodita e passano attraverso quattro stadi larvali (L1-L4). Se c'è carenza di cibo o un numero elevato di individui, *C. elegans* può entrare in uno stadio larvale L3 alternativo, caratterizzato da basso metabolismo, resistenza agli stress e interruzione dei processi di invecchiamento, nel quale può restare per mesi. Durante lo stadio L4, gli ermafroditi producono lo sperma e depongono le uova, ormai adulti. I maschi possono comunque inseminare gli ermafroditi, i quali utilizzeranno preferenzialmente lo sperma del maschio. La durata media della vita di *C. elegans*, nelle linee mantenute in laboratorio a 20 °C, è di circa 2-3 settimane e il tempo di sviluppo è di 3 giorni. Tale rapido ciclo di sviluppo, unito al fatto che ciascun ermafrodita può depositare

Figura 22.15 Ciclo vitale di *Caenorhabditis elegans*.



fino a 300 uova, permette la produzione in laboratorio di diversi milioni di animali adulti al giorno (fig. 22.15).

Il verme è trasparente, per cui l'anatomia dell'organismo è facilmente analizzabile: oggi, inoltre, l'utilizzo di marcatori fluorescenti *in vivo* permette di visualizzare facilmente processi complessi, quali l'embriogenesi e lo sviluppo di singole cellule neuronali.

C. elegans è un sofisticato animale: nonostante l'ermafrodita adulto abbia soltanto 959 cellule somatiche (tutte visibili al microscopio ottico), queste formano molti tessuti e organi differenti, quali pelle, muscoli, intestino, sistema riproduttivo, ghiandole e un sistema nervoso formato da 302 neuroni. Le strutture neurali includono un complesso di organi sensoriali che permettono al nematode di distinguere gli odori e le diverse temperature e percepire la presenza della luce, nonostante *C. elegans* non abbia occhi.

Oltre ad altri numerosi aspetti, l'interesse dei ricercatori si è focalizzato sullo sviluppo di ognuna delle 959 cellule di *C. elegans*: infatti, queste presentano un destino molto ben definito e costante in tutti gli individui, a differenza di quanto avviene nell'embriogenesi dei mammiferi, nella quale il numero di cellule presenti non è geneticamente definito. Numerosi altri studi su *C. elegans* ne hanno analizzato il sistema nervoso e il pattern di connessioni di tutti i 302 neuroni è stato definito nei minimi dettagli.

22.4 L'anfibio *Xenopus laevis* e il suo successore *Xenopus tropicalis*

Le nozioni di base sullo sviluppo embrionale degli animali superiori riportate in tutti i libri di testo si riferiscono in gran parte allo sviluppo degli anfibii. La scelta di questi animali da parte degli embriologi è dovuta al fatto che essi producono uova relativamente grandi e resistenti che si sviluppano al di fuori del corpo della madre; inoltre, essendo vertebrati, gli anfibii non sono molto lontani filogeneticamente dall'uomo. Tra la fine del secolo XIX fino a circa la metà del

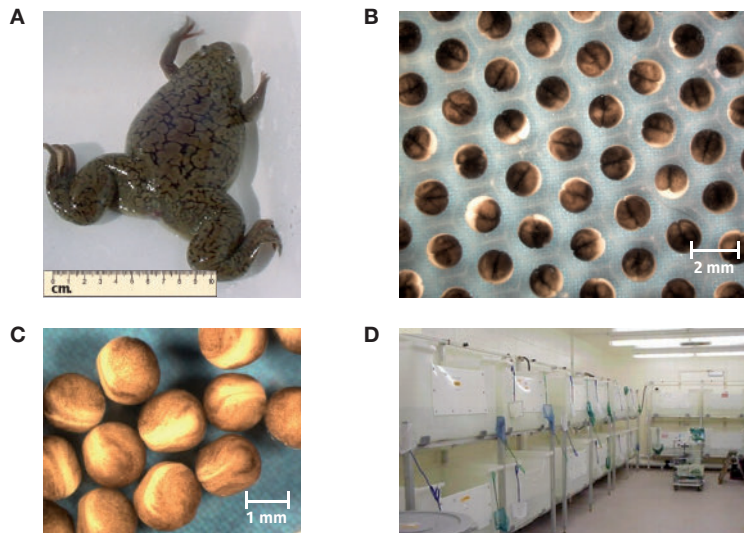


Figura 22.16 L'organismo modello *Xenopus laevis*. (A) Femmina adulta. (B) Segmentazione, stadio a 8 cellule. (C) Embrioni a stadio di neurula. (D) Sistema di vasche per l'allevamento di *Xenopus*.

secolo scorso venivano usati uova, embrioni e individui adulti di varie specie di rane, rospi e tritoni, che venivano reperiti in natura. Uova ed embrioni di queste specie di anfibio sono alla base di studi classici di embriologia come, per esempio, quelli che hanno portato Hans Spemann a scoprire, negli anni venti, il fenomeno dell'“induzione” e successivamente l'“organizzatore primario”. Queste specie però presentavano alcune gravi limitazioni, tra cui la reperibilità in natura delle uova limitata a brevi periodi stagionali e la difficoltà, se non impossibilità, dell'allevamento degli individui adulti in laboratorio.

Intorno alla metà del secolo scorso è stato introdotto per gli studi di embriologia *Xenopus laevis*, che a quell'epoca era già in uso presso i laboratori di alcuni ospedali perché utilizzato per un test di gravidanza (l'iniezione di urina di donna in gravidanza in una femmina di *Xenopus* provoca l'ovulazione). *Xenopus laevis* è un anfibio anuro ma, a differenza delle varie specie dei generi *Rana* o *Bufo*, vive permanentemente nell'acqua e può essere facilmente allevato in acquari in laboratorio (**fig. 22.16**). È possibile ottenere uova mature di *Xenopus* facendo ovulare una femmina mediante iniezione di ormoni o prelevandole chirurgicamente dall'ovaio; embrioni in sviluppo possono essere ottenuti inducendo l'accoppiamento con iniezioni di ormoni nel maschio e nella femmina (**fig. 22.17**). È anche relativamente facile fare una fecondazione *in vitro* mescolando spermatozoi e uova precedentemente prelevati che proseguono poi lo sviluppo in condizioni controllate di laboratorio. A partire dagli anni cinquanta, *Xenopus laevis* è diventato l'organismo d'elezione per studi sullo sviluppo dei vertebrati e, quando negli anni ottanta la Biologia dello sviluppo è arrivata a livello molecolare, *Xenopus* si è dimostrato un organismo modello perfetto, o quasi.

La grandezza dell'uovo (circa 1,3 mm di diametro) e dell'embrione lo rendono ideale per esperimenti di microchirurgia e trapianti di frammenti di tessuti. Questi metodi, accompagnati da analisi a livello molecolare, hanno permesso di identificare importanti molecole segnale e fattori di induzione, quali TGF- β (*Transforming Growth Factor beta*) e FGF (*Fibroblast Growth Factor*), attivine, fattori morfogeni ecc. Si è potuto anche studiare come queste molecole segnale funzionano formando gradienti, che diversi stati di sviluppo corrispondono a diverse combinazioni di fattori di trascrizione e che l'espressione di questi fattori è accesa dai gradienti di molecole segnale. Neanche a dirlo, è poi risultato che tutti questi segnali e le loro interazioni sono molto simili negli altri vertebrati, incluso l'uomo. Abbiamo prima detto che *Xenopus laevis* è un organismo “quasi” perfetto

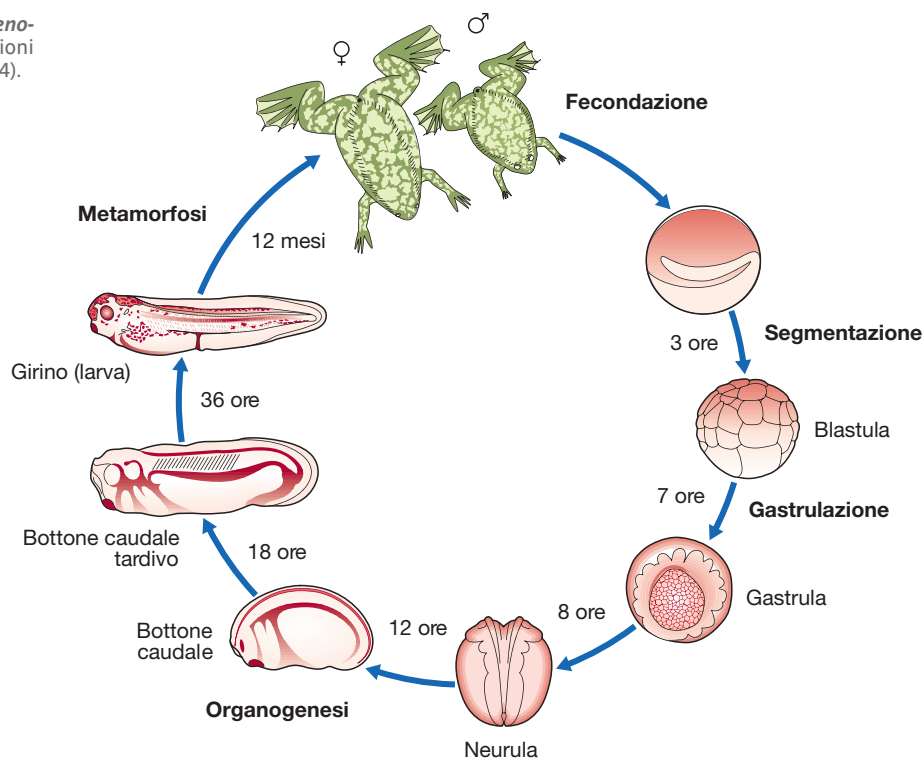
per lo studio dello sviluppo. Il motivo di quel “quasi” è che in effetti qualche difetto ce l’ha. Infatti, come tutti gli anfibî, *Xenopus* è assolutamente inadatto per studi di Genetica per vari motivi:

- il tempo di generazione molto lungo: ci vuole un anno perché una femmina raggiunga la maturità sessuale;
- è difficile ottenere linee pure (*inbred*) di animali;
- sono disponibili pochissimi mutanti;
- la divisione del ciclo vitale in un periodo larvale seguito da una profonda metamorfosi che lo porta allo stadio adulto complica la situazione.

Infine, ci si è resi conto, abbastanza tardi, che *Xenopus laevis* è allotetraploide, cioè il suo patrimonio genetico è il risultato di una duplicazione dell’intero genoma avvenuta circa 30 milioni di anni fa, per cui la maggior parte dei geni è presente in quattro copie: questo fatto rende impossibile molti esperimenti di manipolazione genetica, per esempio fare il knock-out di un gene. A questo i ricercatori hanno in parte ovviato sviluppando metodi per produrre “dominanti-negativi”, cioè inibitori specifici di proteine, oppure forzando l’embrione a fare un eccesso di una particolare proteina iniettandovi il corrispondente mRNA. Inoltre, l’impossibilità di ottenere mutanti è stata in parte superata dalla possibilità di “silenziare” singoli geni nello zigote e durante i vari stadi di sviluppo mediante RNA antisenso o RNAi.

Abbastanza recentemente la comunità scientifica che utilizza *Xenopus* come organismo modello ha deciso di spostarsi dalla specie *Xenopus laevis* alla specie *Xenopus tropicalis*. Quest’ultimo, infatti, ha il vantaggio di essere un normale diploide, quindi con una grandezza del genoma dimezzata, e di avere un tempo di generazione più breve (sei mesi). Molti dei sistemi sperimentali messi a punto in *X. laevis* vengono rimessi a punto per *X. tropicalis* e i geni già clonati e sequenziati per *X. laevis* devono essere riclonati e ristudiati per *X. tropicalis*. È ora in corso

Figura 22.17 Ciclo vitale di *Xenopus laevis*. Disegni degli embrioni tratti da Nieukoop e Faber (1994).



un'iniziativa internazionale per determinare la sequenza del genoma di *Xenopus* e, naturalmente, è stato scelto di farlo su *X. tropicalis*.

Un breve accenno merita la **tecnica di microiniezione in oociti** di *Xenopus*. L'oocita di *Xenopus laevis* è una cellula molto grande (1,3 mm di diametro) con un suo nucleo pure abbastanza grande, tanto che se isolato è visibile a occhio nudo. È possibile, con un microscopio da dissezione e un micromanipolatore relativamente semplice, iniettare materiale (ad es. DNA o RNA o proteine) nel citoplasma o nel nucleo dell'oocita. Per esempio, l'iniezione di un mRNA nel citoplasma dell'oocita risulta in una efficientissima sintesi del corrispondente prodotto proteico, mentre l'iniezione di una proteina potrà stimolare o inibire la trascrizione o la sintesi proteica. Si può dire che l'oocita viene usato un po' come una provetta contenente il macchinario per portare avanti le sintesi cellulari in maniera molto più naturale e, quindi, più efficiente rispetto ai tipici esperimenti *in vitro* con estratti cellulari, citoplasmatici o nucleari. La tecnica di microiniezione in oocita, talvolta definita tecnica "semi-*in vitro*", è utilizzata in molti campi della Biologia anche da ricercatori che indagano sistemi diversi da *Xenopus*.

22.5 Il pesce *Danio rerio* o "zebrafish"

Danio rerio è un piccolo pesce che vive nelle acque dei fiumi tropicali in India e nella parte sud del continente asiatico. È molto popolare negli acquari di tutto il mondo e per le righe orizzontali che caratterizzano il suo aspetto esterno è quasi sempre chiamato con il soprannome di "zebrafish" (**fig. 22.18**).

I pesci zebrafish adulti sono lunghi solo 3-5 cm e possono essere facilmente cresciuti in grande numero in apposite vasche in laboratorio.

Lo sviluppo di zebrafish è molto simile ai meccanismi di embriogenesi nei vertebrati più noti (incluso l'uomo) ma, a differenza di quanto accade nei mammiferi, il suo sviluppo da un uovo fecondato a un organismo adulto avviene all'interno di un uovo trasparente che è deposto esternamente al corpo della madre. Inoltre, gli stessi embrioni sono trasparenti durante i primi giorni di vita. Lo sviluppo è molto rapido: già 24 ore dopo la fecondazione tutti i maggiori organi del pesce sono sviluppati e in tre giorni il pesce esce dall'uovo alla ricerca di cibo (**fig. 22.19**). Dopo 3-4 mesi gli individui sono sessualmente maturi e in grado di generare nuova progenie; una singola femmina riesce a depositare fino a 200 uova alla settimana.

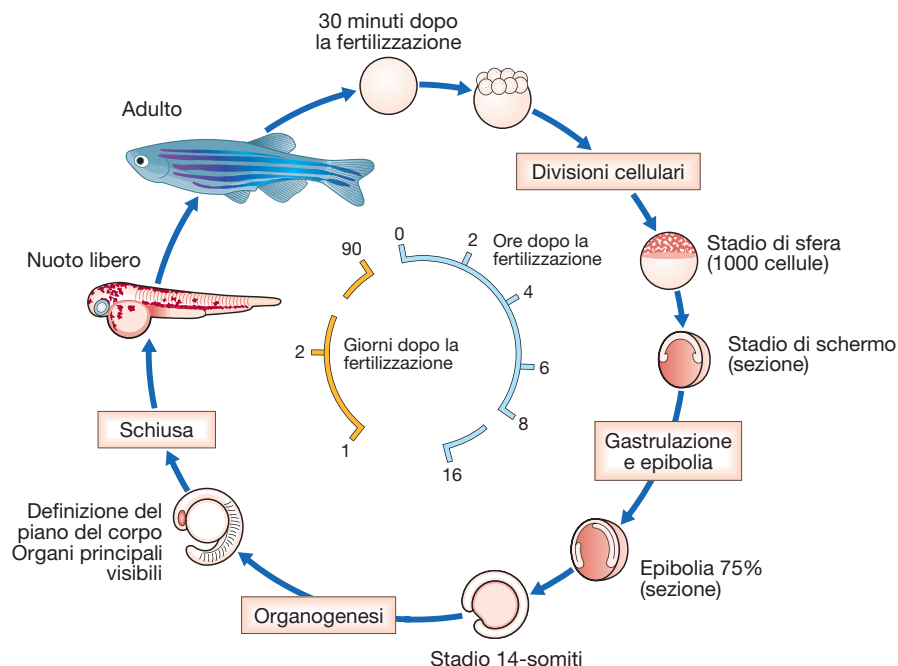
Le caratteristiche fisiologiche indicate sopra hanno reso zebrafish un modello molto utilizzato per lo studio dello sviluppo dei vertebrati. Vediamo ora di analizzare tali caratteristiche un po' più in dettaglio e di valutare in quali ambiti della ricerca biologica zebrafish è oggi maggiormente usato come modello sperimentale.

1. Le piccole dimensioni sia delle larve che dei pesci adulti diminuisce fortemente i costi connessi al mantenimento dei pesci in laboratorio in condizioni facilmente controllabili. Inoltre, nell'ambito di valutazioni dell'impatto ambienta-



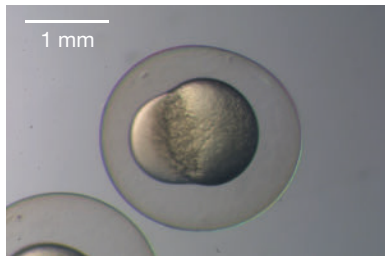
Figura 22.18 Pesce adulto della specie *Danio rerio* o "zebrafish".

Figura 22.19 Ciclo vitale di zebrafish.

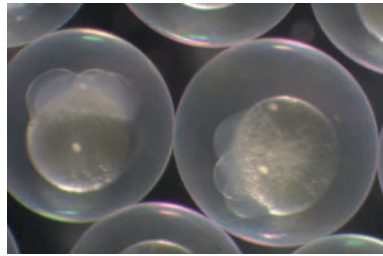


le o di studi tossicologici e farmacologici, i vari test possono essere eseguiti in forma miniaturizzata, diminuendo così i costi collegati alla quantità di sostanze da analizzare e, successivamente, da eliminare. Le piccole dimensioni degli embrioni e dei giovani organismi permettono inoltre lo sviluppo di tecniche di selezione su grande scala (*high-throughput screenings*) che utilizzano piastre di plastica contenenti molti pozzetti di crescita che possono essere replicati in forma anche automatizzata, così da permettere una validazione dei risultati e un'analisi statistica accurata.

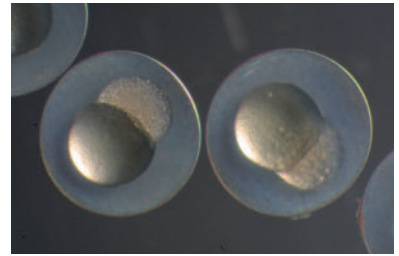
2. Come già accennato, un secondo notevole vantaggio è l'elevata fecondità della specie: una femmina può depositare circa 200 uova a ogni accoppiamento e, a seconda dell'età e delle condizioni ambientali, ciò può ripetersi circa ogni settimana. La rapida maturazione dell'embrione permette di condurre studi su più generazioni, permettendo così di effettuare screening mutazionali per l'isolamento di organismi mutanti. A oggi, due grandi screening sono stati eseguiti in zebrafish per produrre mutanti: uno screening di mutagenesi chimica con un agente mutageno classico, quale l'etil-nitroso-urea, e uno screening per inserzione all'interno di geni: insieme, i due screening hanno permesso la creazione di una collezione di migliaia di mutanti. Recenti sviluppi tecnici hanno anche aperto nuove strade per la valutazione delle funzioni dei geni di zebrafish *in vivo*. In particolare, è molto utilizzata la possibilità di spegnere la funzione di uno specifico gene e di osservare il fenotipo che ne deriva. Ciò si ottiene iniettando nell'uovo fecondato degli oligonucleotidi modificati chimicamente con fosforodiammidato morfolino, in grado di bloccare in modo specifico la traduzione del corrispondente mRNA.
3. Una delle caratteristiche più importanti di zebrafish che lo hanno reso un modello di grande valore per studi di embriogenesi è collegata alla trasparenza dell'uovo e degli embrioni durante i primi giorni di vita, che permette di analizzare e valutare i cambiamenti morfologici al microscopio (fig. 22.20A). La pigmentazione dell'embrione inizia solo dopo 30-72 ore dalla fecondazione e può essere ulteriormente dilazionata mediante opportuni trattamenti dell'embrione *in vivo*. Inoltre, i vari stadi dello sviluppo sono ovviamente più facili da

A

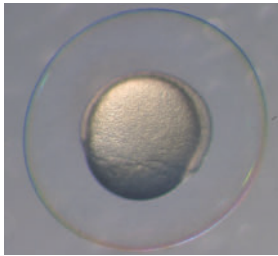
1 cellula"



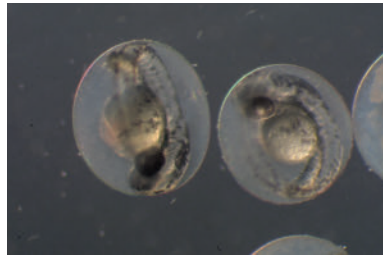
4 blastomeri



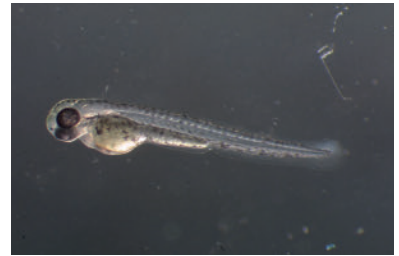
Sfera (oltre 1000 cellule)



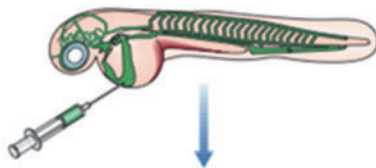
75% epibolia



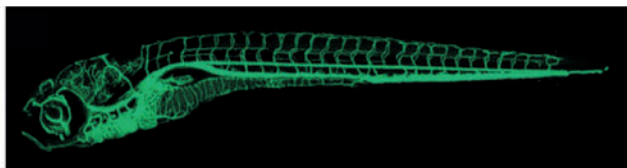
48 ore



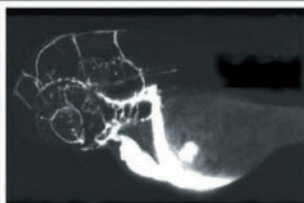
Schiusa (3 gg), nuoto libero

B**Microangiografia**

Circolazione nel pesce wt a 7 giorni



Pesce wt a 2,5 giorni



Pesce m 145 a 2,5 giorni

Figura 22.20 Zebrafish come modello sperimentale. (A) Fotografie di embrioni a diversi stadi di sviluppo. (B) Zebrafish come modello sperimentale per lo studio dello sviluppo del sistema vascolare mediante angiografia, che permette la visualizzazione dei vasi nell'embrione.

analizzare, dato che lo sviluppo dell'uovo e dell'embrione avviene all'esterno del corpo della madre.

4. I mutanti embrionali di zebrafish, anche in presenza di grosse malformazioni morfologiche o funzionali, sono per lo più capaci di sopravvivere al di là del momento in cui l'organo che è alterato nel mutante comincia a svolgere la sua funzione fisiologica nell'organismo sano; questo non accade in embrioni di mammifero, che normalmente muoiono nell'utero della madre.

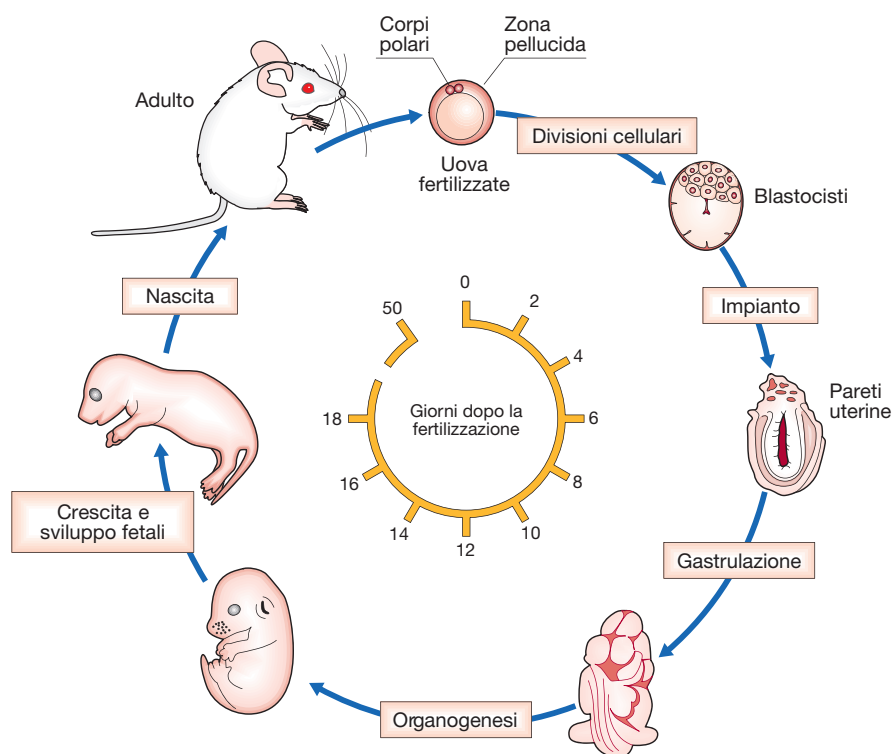
5. Recentemente, l'intero genoma di zebrafish è stato sequenziato. Rispetto a quello umano, esso presenta la complicazione di avere alcune porzioni del genoma duplicate. Ciò si è probabilmente verificato perché durante l'evoluzione dei teleostei l'intero genoma è stato duplicato, un fatto che non è avvenuto nell'evoluzione dei mammiferi. Molti di questi geni duplicati sono però stati persi e solo una piccola parte di essi è oggi presente nel genoma. La conoscenza del genoma di zebrafish ha anche permesso la costruzione di microarray per lo studio dell'espressione genica che sono disponibili commercialmente.

Zebrafish è ed è stato ampiamente utilizzato specialmente in studi di Genetica dello sviluppo. Per le caratteristiche sopra citate, zebrafish è ora molto usato anche in studi farmacologici per lo screening di librerie chimiche, per studi sul meccanismo d'azione di farmaci e per valutazioni di genotossicità. Zebrafish sta diventando sempre più un modello utile per lo studio di malattie dell'uomo. In particolare, in virtù della trasparenza dell'embrione, è diventato il modello d'elezione per lo studio della formazione del sistema vascolare e delle malattie ad esso connesse (**fig. 22.20B**).

22.6 Il topo comune *Mus musculus*

Il ciclo vitale del topo è lento e complicato se confrontato con quello di *D. melanogaster* o *C. elegans*. La gestazione e lo sviluppo embrionale del topo durano circa tre settimane e sono necessarie altre 5-6 settimane affinché i topi neonati raggiungano la pubertà, portando il ciclo vitale complessivo a 8-9 settimane (**fig. 22.21**). Nonostante ciò, il topo è un mammifero ed è evolutivamente molto più vicino all'uomo di un lievito, di un moscerino o di un piccolo nematode. Ovviamente, altri mammiferi, come le scimmie o i primati superiori, sono ancora più simili all'uomo; al di là dei problemi etici, a tali animali non sono però applica-

Figura 22.21 Ciclo vitale del topo comune *Mus musculus*.



bili le varie manipolazioni sperimentali che i topi permettono. Il topo è, quindi, l'organismo d'elezione per trasferire e verificare le conoscenze di base acquisite in organismi modello più semplici e per costruire dei modelli murini estrapolabili quasi totalmente alla specie umana.

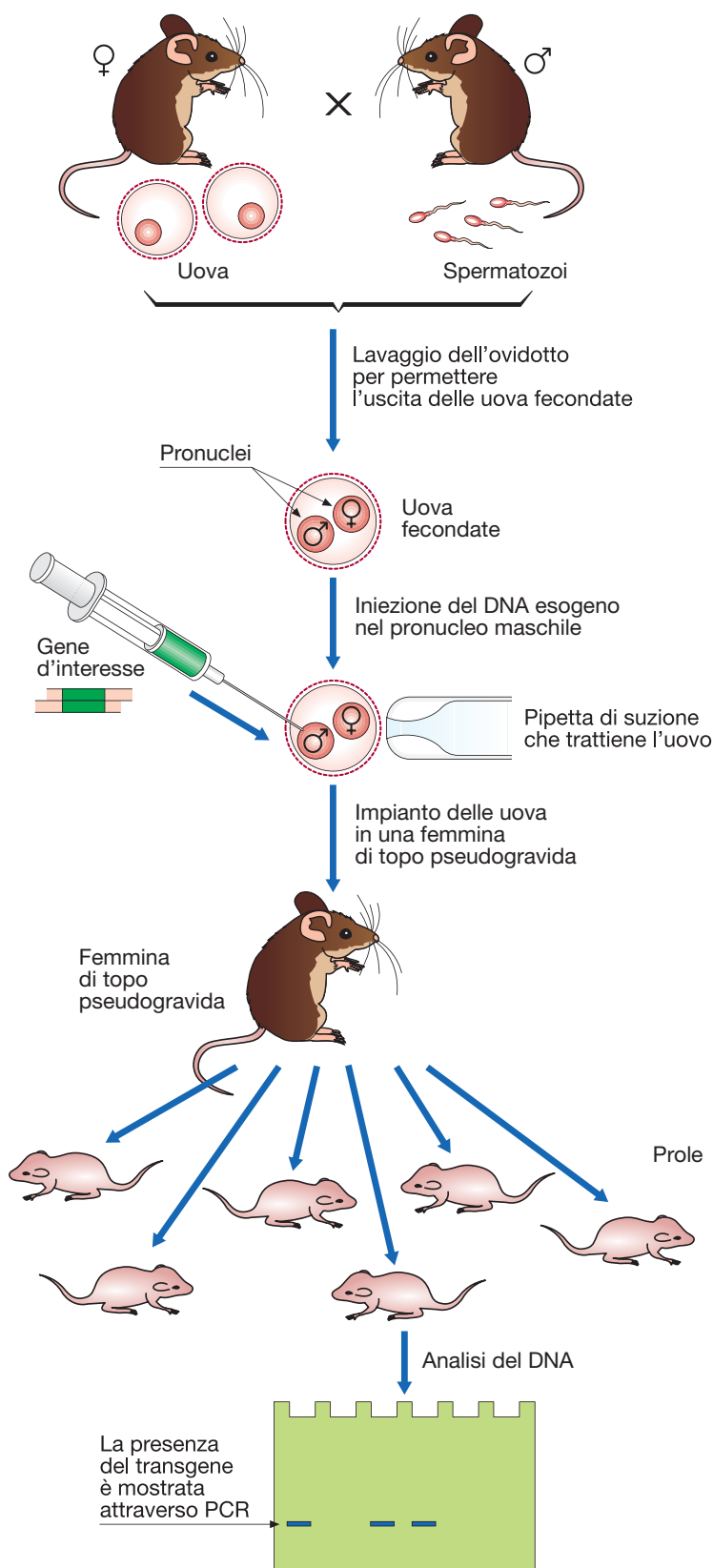
Il corredo cromosomico del topo è molto simile a quello dell'uomo. Possiede 19 coppie di autosomi e due cromosomi sessuali X e Y. La sequenza del genoma del topo è stata pubblicata nel 2002. Anche se la storia evolutiva di uomini e topi si è separata circa 80 milioni di anni fa, i due genomi hanno moltissimo in comune. Il topo ha praticamente lo stesso numero di geni dell'uomo, circa 25 000, ed esiste una perfetta corrispondenza tra uomo e topo per circa l'85% dei geni. Inoltre, c'è un'estesa **sintenia** tra uomo e topo: con tale termine si indica che ampie porzioni di DNA sui cromosomi di uomo o topo contengono le stesse combinazioni di geni e nello stesso ordine. L'analisi genomica comparativa ha, quindi, confermato quanto atteso già da molto tempo: il topo è un modello sperimentale eccellente per lo studio della funzione dei geni umani e dei meccanismi che correlano alterazioni nei geni a malattie dell'uomo.

Le uova di topo sono piuttosto piccole, avendo un diametro di circa 100 micron, analogo alla dimensione della cellula uovo umana. Le divisioni iniziali dell'uovo fecondato sono piuttosto lente: una ogni 12-24 ore. La prima ovvia diversificazione dei vari tipi cellulari avviene allo stadio di 16 cellule, detto *morula*. Le cellule localizzate nelle regioni esterne vanno a formare la placenta, mentre le cellule più interne formano l'ICM (*Inner Cell Mass*). Allo stadio di 64 cellule ci sono solo 13 cellule ICM, ma queste sono in grado di duplicarsi e differenziarsi sino a formare tutti i tessuti del topo adulto. L'ICM è la fonte primaria delle cellule staminali embrionali: tali cellule possono essere fatte crescere in coltura e indotte a formare qualsiasi tipo di cellula mediante l'aggiunta nel terreno di coltura di opportuni fattori di crescita. Allo stadio di 64 cellule (3-4 giorni dopo la fecondazione dell'uovo), l'embrione è chiamato **blastocisti** e, a tale stadio, è pronto per essere eventualmente impiantato nell'utero di una femmina di topo pseudogravida che è stata incrociata con un maschio vasectomizzato così da far scattare tutti gli stimoli ormonali richiesti per l'impianto dell'embrione nell'utero. L'interazione della blastocisti con la parete uterina porta alla formazione della placenta, dopo di che l'embrione entra nello stadio di **gastrulazione**, in cui le cellule dell'ICM formano i tre strati germinali dell'endoderma, mesoderma ed ectoderma. Successivamente, si forma il feto dotato di cervello, spina dorsale e organi interni.

È facile introdurre DNA esogeno nell'embrione di topo: il DNA può essere direttamente iniettato nei pronuclei di uova di topo appena fecondate, analogamente a quanto descritto sopra per l'iniezione di DNA in oociti di *Xenopus*. L'iniezione di un gene d'interesse di cui si vuole analizzare l'espressione in topo viene di solito eseguita nel pronucleo maschile, che è più grande, mentre l'uovo è tenuto fermo da una pipetta di suzione (**fig. 22.22**). Il DNA iniettato può integrarsi nel genoma del pronucleo e, dopo la fusione con il pronucleo femminile, entrerà a far parte del patrimonio genetico dello zigote. Gli embrioni iniettati sono coltivati *in vitro* per un certo periodo di tempo e poi impiantati in una femmina pseudogravida. Se il DNA esogeno si è inserito con successo nel genoma, il topo che nascerà sarà eterozigote per il gene esogeno iniettato. La presenza del transgene nei topi di prima generazione viene verificata prelevando un piccolo campione della coda dei topini per analizzare il loro DNA tramite PCR (*vedi cap. 20*). L'incrocio successivo tra due topi eterozigoti darà origine a una progenie in cui un quarto degli individui sarà omozigote per il transgene.

Tale procedura è stata ampiamente utilizzata e presenta sia svantaggi che vantaggi. Il primo vantaggio consiste nel fatto che il frammento di DNA contenente il gene d'interesse può essere iniettato nel pronucleo senza clonaggio preliminare

Figura 22.22 Produzione di topi transgenici mediante microiniezione. Le uova fecondate sono raccolte lavando l'ovidotto di femmine accoppiate. Il DNA d'interesse è di solito iniettato nel pronucleo maschile e le uova microiniettate sono impiantate in femmine pseudogravide. Tre settimane dopo la nascita, ai topini viene prelevato un piccolo pezzetto della coda su cui viene valutata la presenza del transgene tramite analisi Southern o PCR.



in un vettore opportuno. Inoltre, più copie dello stesso gene possono integrarsi nel genoma in posizioni diverse e più o meno casuali. Gli svantaggi dell'iniezione nel pronucleo sono, però, numerosi:

- geni possono essere introdotti, ma non eliminati, come, invece, si vuole fare se si vuole analizzare la funzione di uno specifico gene di topo (knock-out), come vedremo più avanti;
- l'espressione del gene d'interesse è fortemente influenzata dal contesto cromosomico in cui avviene l'integrazione casuale nel genoma;
- il livello d'espressione del transgene varia nelle generazioni successive essendo influenzato da modificazioni epigenetiche;
- topi transgenici prodotti mediante microiniezione possono dare origine a topi "a mosaico" in cui l'espressione del transgene è limitata solo ad alcuni tessuti e organi, come atteso se l'integrazione del transgene nel genoma avviene dopo la prima divisione cellulare.

Per ovviare a questi inconvenienti e, soprattutto, per poter selezionare le cellule contenenti il transgene e dirigere l'integrazione al locus cromosomico desiderato, sono state sviluppate nuove tecnologie e vettori e sono normalmente utilizzate **cellule staminali embrionali (ES)**. Le cellule ES sono cellule indifferenziate che vengono isolate dalla blastocisti (**fig. 22.23**). Possono essere coltivate *in vitro* e mantenute in coltura per molte generazioni. Le cellule ES hanno la potenzialità di dare origine a tutti i tipi cellulari presenti in un animale adulto, inclusi i gameti. Sono state sviluppate anche delle procedure che permettono di ottenere il differenziamento delle cellule ES in uno specifico tipo cellulare.

Queste proprietà rendono le cellule ES particolarmente utili per la produzione di topi transgenici. Il vantaggio più importante delle cellule ES consiste nel fatto che il transgene da utilizzare può essere manipolato *in vitro* e le cellule ES contenenti il transgene possono essere selezionate prima di essere iniettate nella blastocisti (**fig. 22.23**). Le cellule contenenti il transgene si uniranno alla massa delle altre cellule della blastocisti e prenderanno parte alla formazione dei vari tessuti del futuro topo, che sarà, quindi, un topo chimerico. Di solito le cellule ES manipolate *in vitro* derivano da topi con un colore del pelo diverso da quelle del topo nella cui blastocisti vengono iniettate le cellule ES manipolate, così da verificare il loro contributo guardando semplicemente il colore del pelo (**fig. 22.23**). Gli animali chimerici verranno poi incrociati con animali selvatici in modo da ottenere dei veri eterozigoti che, incrociati tra di loro, daranno origine a topi omozigoti. Partendo da cellule ES sono quindi necessari due incroci per ottenere degli omozigoti per il transgene.

Il vantaggio più importante nell'utilizzo delle cellule ES è la loro efficienza nei meccanismi di ricombinazione omologa. Ciò permette di dirigere l'integrazione del transgene dando origine a topi in cui un gene è completamente inattivato (**topi knock-out**) o in cui la copia selvatica del gene è sostituita da una copia opportunamente modificata (**topi knock-in**). Lo sviluppo di tecniche e vettori che hanno facilitato la ricombinazione omologa di transgeni in loci specifici sui cromosomi di topo ha portato all'assegnazione nel 2007 del premio Nobel per la medicina a Mario Capecchi, Oliver Smithies e Martin Evans.

Nonostante le cellule ES siano capaci di ricombinazione omologa, possono avvenire anche eventi di ricombinazione non omologa, per cui è necessario saper distinguere e selezionare l'evento desiderato. Una strategia frequentemente utilizzata per ottenere il knock-out di un gene attraverso meccanismi di ricombinazione omologa è descritta in **figura 22.24**. Tale metodo utilizza un frammento di DNA in cui, per esempio, nel terzo esone del gene che si vuole inattivare è stato clonato un marcatore di selezione quale il gene *neo*, che, conferendo la resistenza alla neomicina, permette alle cellule che lo contengono di crescere in presenza di un

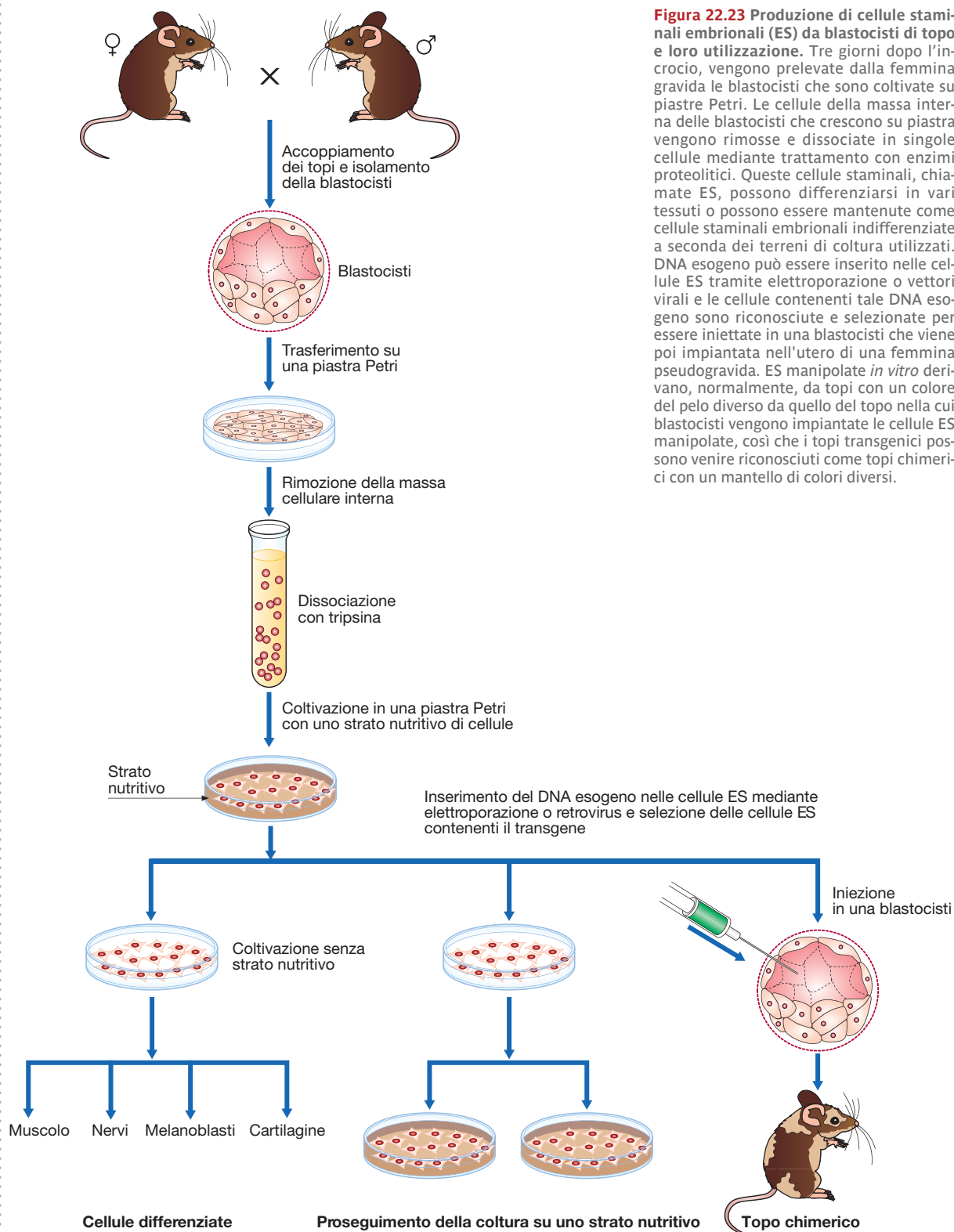


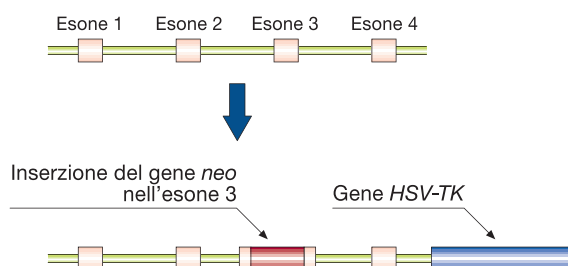
Figura 22.23 Produzione di cellule staminali embrionali (ES) da blastocisti di topo e loro utilizzazione. Tre giorni dopo l'incrocio, vengono prelevate dalla femmina gravida le blastocisti che sono coltivate su piastre Petri. Le cellule della massa interna delle blastocisti che crescono su piastra vengono rimosse e dissociate in singole cellule mediante trattamento con enzimi proteolitici. Queste cellule staminali, chiamate ES, possono differenziarsi in vari tessuti o possono essere mantenute come cellule staminali embrionali indifferenziate a seconda dei terreni di coltura utilizzati. DNA esogeno può essere inserito nelle cellule ES tramite elettroporazione o vettori virali e le cellule contenenti tale DNA esogeno sono riconosciute e selezionate per essere iniettate in una blastocisti che viene poi impiantata nell'utero di una femmina pseudogravida. ES manipolate *in vitro* derivano, normalmente, da topi con un colore del pelo diverso da quello del topo nella cui blastocisti vengono impiantate le cellule ES manipolate, così che i topi transgenici possono venire riconosciuti come topi chimerici con un mantello di colori diversi.

antibiotico noto come G418. Il vettore contiene inoltre il gene *HSV-TK* che codifica per il gene codificante la timidino-chinasi del virus Herpes Simplex (HSV), il cui prodotto genico conferisce sensibilità delle cellule al farmaco antivirale ganciclovir. Un doppio evento di ricombinazione in regioni fiancheggianti l'esone che porta il marcatore di selezione *neo* determina l'inattivazione del gene a causa della grossa inserzione presente nella regione codificante il terzo esone. Le cellule in cui è avvenuto tale evento di ricombinazione omologa possono venire specificamente selezionate in quanto sono resistenti sia al G418, perché è stato inserito il gene per la resistenza, sia al ganciclovir, perché non è stato inserito il gene per la sensibilità. Se il costrutto originario si è integrato casualmente nel genoma non tramite ricombinazione omologa ma tramite ricombinazione illegittima, l'intero costrutto si sarà inserito da qualche parte nel genoma, dando origine a cellule che sono resistenti a G418, ma sensibili a ganciclovir. È in questo modo possibile riconoscere specificatamente le cellule ES in cui è avvenuta la ricombinazione omologa e il conseguente knock-out del gene d'interesse. Tale situazione dovrà, ovviamente, essere opportunamente controllata tramite PCR o Southern blot.

La possibilità di generare topi knock-out (ma anche knock-in) di specifici geni ha grandi potenzialità per comprendere la funzione di quel gene nell'animale. I topi knock-out possono avere tre fenotipi:

- un fenotipo letale che porta alla morte dell'embrione a stadi più o meno precoci dello sviluppo;
- un fenotipo mutante che non pregiudica la vitalità dell'embrione, ma mostra un fenotipo analizzabile nell'organismo dopo la nascita;
- un fenotipo normale, in quanto l'inattivazione del gene non causa alcun fenotipo evidente. In quest'ultimo caso non si può però facilmente escludere che il topo knock-out presenti dei sottili fenotipi non facilmente visibili e che riguardino aspetti collegati al comportamento o alla risposta a particolari condizioni non facilmente prevedibili.

Il gene di tipo selvatico in esame è modificato *in vitro*



L'esone 3 contenente il gene *neo* è inserito nel genoma tramite ricombinazione, con conseguente separazione del gene *TK*

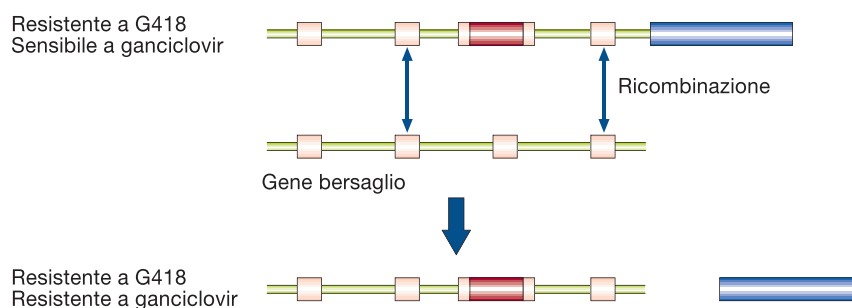


Figura 22.24 Utilizzo di vari marcatori per la selezione di topi knock-out che derivano da eventi di ricombinazione omologa. Come descritto nel testo, un transgene contenente il gene *neo* in un esone del gene che si vuole inattivare e il gene *HSV-TK* a valle del costrutto, può essere selezionato in quanto resistente all'antibiotico G418 e al ganciclovir.

Un fenotipo letale riflette, di solito, una funzione molto rilevante del gene la cui alterazione determina una morte precoce dell'embrione. Poiché il topo è un organismo diploide, i mutanti che rientrano in questa classe possono essere analizzati allo stato eterozigote per valutare se la presenza di una sola copia del gene determini un qualche fenotipo. Alternativamente, è diventato di uso comune la produzione di knock-out condizionali che possono essere prodotti solo in particolari condizioni.

In esperimenti di knock-in, invece, il gene selvatico viene spesso sostituito da una forma mutata del gene modificata in una regione specifica in seguito a procedure di mutagenesi sito-specifica (*vedi* cap. 20). Ciò permette di analizzare la funzione di specifiche regioni di una proteina. È anche possibile sovrapprodurre la proteina d'interesse per valutare l'effetto di variazioni nei livelli di proteina.

L'utilizzo di topi knock-out e knock-in è stato negli ultimi anni di estrema importanza per capire le funzioni di geni di mammifero coinvolti in tutti gli aspetti della loro biologia e per costruire modelli murini di malattie umane di grande potenzialità e utilità per lo sviluppo di nuove terapie e per la scoperta di nuovi farmaci. La somiglianza fenotipica tra pazienti umani con una specifica malattia e il corrispondente modello murino è spesso veramente sorprendente.

22.7 La pianta *Arabidopsis thaliana*

Arabidopsis è una piccola pianta erbacea di altezza compresa tra i 10 e 50 cm (**fig. 22.25**). Ha fusto eretto coperto da peli e due tipi di foglie: quelle basali più grandi e quelle disposte sul fusto più piccole. I fiori hanno petali bianchi, si presentano a grappoli e sono ermafroditi, presentando gli attributi di entrambi i sessi. *Arabidopsis* è oggi diffusa in tutte le zone a clima temperato ed è spesso considerata una pianta infestante in piantagioni d'interesse agronomico.

Nonostante ciò, *Arabidopsis* è l'organismo modello più comunemente utilizzato in studi di Genetica e Biologia Molecolare delle piante: basti pensare che,

Figura 22.25 Fotografia di una piantina di *Arabidopsis thaliana*.



mentre negli anni settanta erano pochi i laboratori che utilizzavano *Arabidopsis*, nel 2007 è stato calcolato che oltre 6000 laboratori e 16 000 ricercatori utilizzavano tale organismo come modello. C'è da domandarsi il perché di tale scelta.

Dal punto di vista economico le piccole dimensioni di *Arabidopsis* lo rendono un modello ideale e poco costoso per gli spazi ristretti dei laboratori o delle serre connessi agli istituti di ricerca che si occupano di biologia delle piante. Inoltre, ha un ciclo vitale molto breve: trascorrono circa sei settimane dalla germinazione di una piantina fino a che questa dia origine a una pianta adulta in grado di produrre i propri semi. Il breve ciclo vitale riduce notevolmente i tempi della ricerca, che quando si lavora con le piante sono molto lunghi e fortemente condizionati dalle condizioni ambientali naturali.

Il primo mutante di *Arabidopsis*, noto come *Agamous*, è stato prodotto sin dal 1873, anche se il gene corrispondente è stato clonato solo nel 1990. La disponibilità di una collezione di mutanti generati negli anni cinquanta tramite mutagenesi indotta con raggi X è stato un altro motivo che ha favorito *Arabidopsis* come sistema modello. Tuttavia, il fatto che ha reso questo organismo estremamente popolare tra i genetisti e i biologi molecolari delle piante è stata la pubblicazione nel 1987 di una procedura basata sull'utilizzo di *Agrobacterium tumefaciens* (vedi cap. 20) per trasformare i semi della pianta, permettendo, così, un approccio sperimentale non basato su colture di tessuto vegetale.

L'utilizzo della mutagenesi inserzionale tramite trasposoni ha poi portato alla produzione sistematica di mutanti in tutti i geni dell'organismo e il passo finale nell'accettazione di *Arabidopsis* come modello del mondo vegetale è stato il sequenziamento del suo genoma, la cui prima bozza è stata pubblicata nel 2000 e da allora più volte modificata e corretta.

Arabidopsis ha cinque cromosomi, ma quello che è più importante è che le dimensioni del suo genoma sono contenute rispetto alle dimensioni di altre specie vegetali. Infatti, il genoma di *Arabidopsis* è di soli $1,25 \times 10^8$ pb e, dal 2000 in poi, lo scopo dei ricercatori è stato di comprendere la funzione dei circa 29 000 geni di questa specie vegetale mediante l'analisi di mutanti e la produzione di ceppi knock-out, analogamente a quanto abbiamo già ampiamente descritto per i modelli animali.

Lecture di approfondimento

- Antoshechkin, I. e Sternberg, P.W. (2007), «The versatile worm: genetic and genomic resources for *C. elegans* research», *Nature Reviews Genetics*, **8**, pp. 518-532.
- Beller, M. e Oliver, B. (2006), «One hundred years of high-throughput *Drosophila* research», *Chromosome Research*, **14**, pp. 349-362.
- Boone, C., Bussey, H. e Andrews, B.J. (2007) «Exploring genetic interactions and networks with yeast», *Nat. Rev. Genet.*, **8**, pp. 437-449.
- Ingham, P.W. (2009), «The power of the zebrafish for disease analysis», *Human Molecular Genetics*, **18**, pp. 107-112.
- Philpott, A. e Yew, P.R. (2008), «The *Xenopus* cell cycle: an overview», *Molecular Biotechnology*, **39**, pp. 9-19.

Indice analitico

A

α -amanitina, 290
 α -complementazione, 522
 α -CTD, 237, 241, 262
AAUAAA, 342
accartocciamento, 239
accession number, 604
acetilazione, 488
acetil-CoA, 488
ACS (*ARS Consensus Sequence*), 144
adattatori (*adapters*), 541
AFM (*Atomic Force Microscopy*), 580
alberi di Natale (geni rRNA), 92, 330
albero filogenetico, 606
allolattosio, 255
allungamento della traduzione
 negli eucarioti, 410, 415
 nei procarioti, 412
alphoid DNA o satelliti α , 113
amber, mutazione non senso, 416
amminoacil-tRNA sintetasi, 389, 392
amplificazione
 dei geni per l'rRNA, 180
 genica, 180,
animali transgenici, 531
annotazione di sequenze proteiche, 603
anse cromosomiche, 89, 456
anticodone, 64
antigene T, 165
antiterminazione, 248, 268
APE endonucleasi, 192
apoptosi, 200
appaiamenti di Hoogsteen, 29

apparato basale (trascrizione di Pol II), 294
aptameri, 51
Arabidopsis thaliana, 362, 370, 634
Archeobatteri, 97
ARE (*AU-Rich Element*), 433
Argonauta (Ago) 467
ARS (*Autonomously Replicating Sequence*) di lievito, 143
asco, 208, 611
ASH1, *vedi* localizzazione degli mRNA
asimmetria composizionale, 100
assorbanza A, 498
ATM, proteine chinasi, 201
ATR, proteine chinasi, 201
att, siti, 209
attenuazione, 265
 peptide leader, 268
 sequenza leader, 268
attivatori (della trascrizione), 251
autosplicing, 349
avvolgimento
 plectonemico 36
 toroidale 36

B

β -galattosidasi, 254, 255, 522
BAC (*Bacterial Artificial Chromosome*), 592
banche genomiche, *vedi* librerie genomiche
banche di cDNA, *vedi* librerie di cDNA
batteriofago, *vedi* fago
BBP (*Branch point Binding Protein*), 353
BER (*Base Excision Repair*), riparazione per escissioni di
 basi, 188, 190
 long-patch BER, 192

short-patch BER, 192
 BHLH (*Basic Helix-Loop-Helix*), 316
 bilancio energetico della traduzione, 416
 BLAST (*Basic Local Alignment Search Tool*), 599
 blastocisti, 629
 bolla di trascrizione, 231
 bomba di Avery, 9
 BP (*Branching Point*), *vedi* ramificazione, sito di *branch migration*, *vedi* migrazione della giunzione
 BRE (*B Responsive Element*), 296
 bruco, modello per inizio trascrizione, 239
 bromodominio, 95
 browser genomici, 604

C

C. elegans (*Caenorhabditis elegans*), 620
 CaMV (virus del mosaico del cavolfiore), 535
 CAP (proteina attivatrice dei geni catabolici), 260
 cap (cappuccio), 299, 406
 cap binding protein, *vedi* eIF4E
 capping (aggiunta del cap), 339, 342
 carioferine, 376
 Cdc25, 202
 Cdc6, 177
 CDE-I, CDE-II e CDE-III, 80
 CDK (*Cyclin-Dependent Kinase*), 175, 176, 179, 203, 485
 Cdt1, 177
 CEN, regioni centromeriche, 80
 CENH3, proteina centromerica, 218
 CENP-B, proteina del cinetocore, 218
 Centrifugazione

- in gradiente di CsCl, 510
- in gradiente di saccarosio, 507
- isopicnica, *vedi* centrifugazione in gradiente di CsCl

 centro della peptidil transferasi del ribosoma, 388
 centro di decodificazione del ribosoma, 388
 centromeri, 80-81
 ceppi eterotallici, di lievito, 611
 ceRNA, 470
 CFI (*Cleavage Factor I*), 343
 CFII (*Cleavage Factor II*), 343
 CGS (*Clone-by-clone shotgun Genome Sequencing*), 592
 chaperoni molecolari, 237, 477
 checkpoint, 175

- da danno al DNA, 197

 chiasmi, 77, 204
 chinasi a doppia specificità, 484
 chinasi ciclina-dipendenti, 175, 485
 ChIP (*Chromatin ImmunoPrecipitation*), 179, 579
 ChIP-Seq, 602
 “chiusura lampo” di ribosio (*ribose zipper domain*), 50

ciclina B mRNA, 429
 cicline, 175, 485
 ciclo cellulare (fasi G1, S, G2, M), 173
 ciclo dello spliceosoma, 353
 cinetocore, 80
 CKI (*Cyclin-dependent Kinase Inhibitors*), 176, 487
 classe eterogenea di RNA nucleari, *vedi* hnRNA
 cleavage factor I, *vedi* CFI
 clonaggio di un segmento di DNA, 513
 clonazione animale

- clonazione in *Xenopus*, 180
- pecora Dolly, 180

 CLUSTAL (programma bioinformatico), 599
 coda di poli(A), funzioni, 345
 codice degli istoni, 488
 codice genetico, 55-68

- “congelato”, 66
- decifrazione, 56-60
- degenerato, 64
- mitocondriale, 67, 123
- non ambiguo, 64
- origine ed evoluzione, 65
- universale, 63, 126

 codone (o tripletta), 57

- di stop (di terminazione o non senso), 60, 63
- sinonimi, 64

 co-immunoprecipitazione, 584
 co-integrato, 216
 colinearità gene-proteica, 55
 collana di perle (o fibra cromatinica da 10 nm), 87
 co-localizzazione, 584
 complementazione *in vitro*, 156
 complessi di rimodellamento della cromatina, 457
 complesso aperto, inizio della trascrizione, 234
 complesso chiuso, inizio della trascrizione, 234
 complesso di pre-inizio 43S, della traduzione, 404, 406
 complesso di pre-innesco, della sintesi del DNA, 158
 complesso innesco-stampo (*primer-template complex*), 150
 condensazione dei genomi fagici, 74
 confronto e allineamento di sequenze, 598
 coniugazione, meccanismo di scambio genetico, 99
 controlli traduzionali autogeni, 426
 controllo qualità dell'mRNA, 435
 conversione genica, 122, 208, 209
 co-purificazione, 584
 core promoter, *vedi* promotore minimo
 corepressore, 265
 corpi d'inclusione, 525
 correzione di bozze (proofreading)

- nella sintesi del DNA, 153
- nella traduzione, 394

 co-sedimentazione, 584
 cosmidi, 529

Cot, 499, 501
 Cot^{1/2}, 501
 CPE (*Cytoplasmic Poly(A)denylation Element*), 345, 429
 CPEB (*CPE Binding protein*), 429
 CpG, 450, 452, 454
 CPSF (*Cleavage and Poly(A)denylation Specificity Factor*), 342
 criptogeni, 372
 cristallografia a raggi X, 48
 CRM, retrotrasposone delle piante, 217
 cromatina, struttura, 84-91
 sensibilità alle nucleasi, 93
 siti ipersensibili alle endonucleasi, 93
 cromodominio, 95
 cromomero, 78
 cromosomi
 a spazzola (o piumosi), 78
 politénici, 78, 618
 cropping dei microRNA, 466
 crossing over, 204
 diseguale, 122
 CRP, *vedi* CAP
 cruciforme, struttura, 30
 CstF (*Cleavage stimulation Factor*), 342
 CTD (*Carboxy-Terminal Domain*), 291, 299, 341

D

Danio rerio (zebrafish), 625
 dbEST (banca dati di sequenze EST), 595
 DDBJ (banca dati biomolecolari giapponese), 604
 deadenilazione dell'mRNA, 433
 debranching, *vedi* deramificazione
 decapping, enzimi di, 435
 degradazione dell'mRNA, 435
 degradosoma, 432
 demetilasi, 452, 490
 denaturazione del DNA, 497, 498, 547
 densità di galleggiamento (*buoyant density*), 510
 deossiribosio, 12
 deramificazione, enzima di, 352
 determinazione (nello sviluppo di *Drosophila*), 617
 determinazione del sesso in *Drosophila*, 367
 DICER, 334, 465
 dicing dei microRNA, 466
 diffrazione dei raggi X, 18
 digestione parziale del DNA, 538
 diminuzione
 cromatinica, 182
 cromosomica, 182
 genica, 182
 diskerina, 339

dito B (B-finger), dominio di TFIIB, 298
 DNA, struttura, 17-25
 a "sequenza semplice", 112
 base stacking, 24
 curvo, 32
 forma A, 27
 forma B, 26
 forma Z, 27
 mediamente ripetitivo, 113, 502
 quartetto di G, 29, 83
 solco maggiore, 23
 solco minore, 23
 stabilità della doppia elica, 23
 struttura chimica, 12
 struttura fisica, 17
 strutture alternative, 25
 topologia, 33
 tripla elica, 28
 unico (non ripetitivo), 502
 DNA glicosilasi, 191
 DNA ligasi, 154, 192, 516, 521, 541
 DNA polimerasi, 150
 polimerasi I, 151, 162
 polimerasi II, 151
 polimerasi III, 160
 polimerasi α -primasi, 167
 polimerasi δ e ϵ , 167
 DNA primasi, 154, 160
 di *Escherichia coli*, 160
 DNA topoisomerasi, 33, 40, 42, 158, 245
 DnaA, 157
 DnaB, 158
 DnaC, 158
 DNA satellite, 111, 511
 DNasi (deossiribonucleasi), 330
 DNMT (DNA metil transferasi), 452
 dNTP, 150
 dogma centrale, 55-56
 domini di legame al DNA, 311, 312
 BHLH (*basic helix-loop-helix*), 316
 cerniera di leucine (*leucine zipper*), 314
 dita di zinco (*zinc finger*), 294, 567
 HD (omeodominio), 312
 HLH (motivo elica-ansa-elica), 315
 HTH (motivo elica-giro-elica), 262, 312
 HTH winged, 312
 dominio A, *vedi* ACS
 doppio ibrido, saggio del, 310
 DROSHA, 465
Drosophila melanogaster, 78, 140, 538, 616
 DSB (*Double Strand Break*), 169, 197
 DUE (*DNA Unwinding Element*), 144
 duplicazioni segmentali, 119

E

4E-BP (*4E Binding Protein*), 425
 E1 (*ubiquitin activating Enzyme*), 479
 EBI (European Bioinformatics Institute), 589, 604
 E-box (sequenza di alcuni promotori), 316
 editing dell'RNA, 368
 conversione di una base in un'altra, 368, 369
 inserzione o delezione di nucleotidi, 372
 eEF (*eukaryotic Elongation Factor*), traduzione, 415
 EF (*Elongation Factor*), traduzione nei procarioti, 411
 EF-G, 411
 EF-Ts, 411
 EF-Tu, 411
 effetto ipercromico, 498
 effetto letale sintetico, 582
 eIF (*eukaryotic Initiation Factor*) per la traduzione, 406
 eIF2, 406
 eIF2/GTP, 406
 eIF3, 406
 eIF4A, 406
 eIF4B, 406
 eIF4E, 342, 406, 425, 429
 eIF4F, 406
 eIF4G, 406
 eIF5B, 406, 408
 eIF6 (fattore antiassociativo), 406, 408
 EJC (*Exon Junction Complex*), 355, 437
 elementi LTR, 116
 elementi mobili del mais, 213
 elementi non-LTR, 116
 elementi prossimali dei promotori di Pol II, 296
 elementi trasponibili, *vedi* trasposoni
 elemento -10 dei promotori procariotici, 255
 elemento -35 dei promotori procariotici, 255
 elemento UP dei promotori procariotici, 237, 264
 elettroforesi su gel, 516
 elettroporazione, 521
 elica R (elica di riconoscimento, del motivo HTH), 277, 312
 EMBL data library, 604
 EMSA (*Electrophoretic Mobility Shift Assay*) o "gel-shift", 575
 endogenoma, 98
 endonucleasi, 330, 360, 516
 enhanceosoma, 306
 enhancer, 296, 306, 366
 ENSEMBL, 604
 enzimi di restrizione (ER), 516
 eredità epigenetica, 449
 eredità uniparentale, 128
 eRF1, 416
 eRF3, 416
 ES (cellule staminali embrionali), 631
 ESE (*Exonic Splicing Enhancer*), 366

esone, 105, 352
 costitutivo, 365
 facoltativo (o alternativo), 365
 esonizzazione, 108
 esonucleasi, 81, 329, 330
 3'→5', 329, 330, 432
 5'→3', 329, 330, 432
 esperimento
 del frullatore 10
 di Gurdon 180
 di Meselson e Stahl, 134
 espressione costitutiva basale, 251
 ESS (*Exonic Splicing Silencer*), 366
 estratto cellulare, 58
 estremità
 piatte, 518
 sporgenti, 518
 eterocromatina, 76
 costitutiva, 76
 facoltativa, 76
 eucromatina, 76
 exon shuffling, 107, 365

F

fagi trasponibili, 213
 fago, 4
 fago lambda, 271
 cI, 275, 276
 ciclo lisogenico, 271
 ciclo litico, 271
 cII e *cIII*, 274, 283
 cro, 273, 276
 induzione del ciclo litico, 281
 induzione lisogenica, 275
 legame cooperativo del repressore CI al promotore, 281
 operatori O_{R1} , O_{R2} , O_{R3} , 276
 promotori P_R , P_L , P_{RM} , 276, 280
 regione di controllo (o di immunità) 271, 278
 fago Mu, 216, 218
 famiglie geniche, 119-121
 famiglia dei geni delle globine, 120
 fase (modulo) di lettura, 60
 fattore di stringenza, 424
 fattore rho, 268
 fattore RRF (*Ribosome Recycling Factor*), 416
 fattori basali o generali (GTF) di Pol II, 289, 291
 TFIIA, 297
 TFIIB, 296
 TFIID, 297
 TFIIE, 298
 TFIIH, 195, 298
 fattori di regolazione di Pol II, 304

- Ap1 e Ap2, 306
- CREB, 323
- NF-kB, 324
- Oct-1 e oct-2, 306
- Sp1, 306
- fattori α (dell'RNA polimerasi procariotica), 291
 - alternativi, 229
 - alternativi del fago SPO1, 232
 - alternativi di *B. subtilis*, 243
 - alternativi di *E. coli*, 243
- fattori di trascrizione della Pol I
 - SL1, 292
 - UBF (*Upstream Binding Factor*), 292
- fattori di trascrizione della Pol III
 - TFIIIA, 294
 - TFIIIB, 294
 - TFIIIC, 294
- fedeltà di replicazione del DNA, 152
- FEN1 (*Flap EndoNuclease I*), 192
- ferritina (regolazione traduzionale), 428
- fibrillarina, 338
- filamento
 - codificante, 125, 240
 - continuo (leading strand), 154
 - ritardato (lagging strand), 154
 - stampo, 197, 234
- flagellina
 - 1 (H1), 219
 - 2 (H2), 219
- flusso dell'informazione genetica, 56
- fMet-tRNA^{fMet} (tRNA di inizio procariotico), 401, 403
- footprinting, 179, 575
- forcella di replicazione (*replication fork*), 154
- forcina, struttura, 31
- fosfatasi, 483
- frammenti di Okazaki, 154, 162
- FRET (trasferimento di energia di risonanza di fluorescenza), 239, 586
- reporter, 570
- selettori omeotici, 619
- genoma dei cloroplasti, 129-130
- genoma mitocondriale, 123-129
- genome editing, 565
- genomi eucariotici, 102-123
- genomi procariotici, 97-102
- genomi virali, 72, 130
- Genomica funzionale 571, 614
- genoteche, *vedi* librerie genomiche
- GFP (*Green Fluorescent Protein*), 305, 570, 586
- gin (G invertasi), 219
- giunti (linker), 541
- giunzione di Holliday, 205
- glicosilazione di proteine, 490
- glnA, geni controllati da NtrC, 264
- GG-NER (*Global Genome NER*), 193
- granuli citoplasmatici di RNA, 444
- GreB, fattore di allungamento della trascrizione nei procarioti, 244
- gRNA, *vedi* RNA guida
- gruppi di epistasi, 193
- gruppo del fago, 4, 5
- guanilil transferasi, 339

H

- HAT (*Histone Acetyl Transferase*), 95, 457, 489
- HDAC (*Histone Deacetylase*), 95, 457, 489
- HERV (*Human Endogenous RetroVirus*), 118
- Hin (H invertasi), 220
- HMG (*High Mobility Group*), 84, 292
- HML e HMR (loci inattivi fiancheggianti il locus MAT di lievito), 122, 221
- HMT (istone metil transferasi), 95, 457
- hnRNA (*heterogeneous nuclear RNA*), 339, 349
- HO (endonucleasi specifica di MAT), 222
- HR (*Homologous Recombination*), 189, 197, 565

G

- Gal4, 309
- gastrulazione, 629
- GCN4 mRNA, 430
- Gcn5, 489
- Genbank, 604
- gene regolatore, 253
- Genetica inversa, 571, 572, 614
- Geni
 - CDC (*Cell Division Cycle*), 173, 614
 - della segmentazione, 619
 - discontinui (*split genes*), 347
 - housekeeping, 306
 - reiterati, evoluzione parallela (o concertata), 121

I

- ibridazione di acidi nucleici, tecniche, 497
 - a saturazione, 503
 - in situ*, 503
 - su colonia, 506, 544
 - su placca, 506, 545
- IF1 (fattore di inizio 1, per la traduzione nei procarioti), 402
- IF2 (fattore di inizio 2, per la traduzione nei procarioti), 402
- IF3 (fattore di inizio 3, per la traduzione nei procarioti), 402
- impacchettamento del DNA, 71, 73
 - del DNA eucariotico, 75

del genoma batterico, 74
 nei virus filamentosi, 72
 nei virus icosaedrici, 73
 traslocazione, 73
 impalcatura (scaffold) del cromosoma, 89
 impilamento delle basi (*base stacking*), 24
 imprinting genetico, 454
 ingegneria genetica o tecnologie del DNA
 ricombinante, 205, 512
 inibitori delle CDK, 176
 inizio della traduzione cap-dipendente negli eucarioti,
 405
 inizio della traduzione cap-indipendente negli
 eucarioti, 410
 inizio della traduzione nei procarioti, 403
 innesco, o primer, 150
 INR (iniziatore) della trascrizione, 296
 Int (integrasi), 285
 integrasi di λ , 209
 interbande e bande nei cromosomi politenici, 78
 interferenza da RNA, vedi RNAi
 introni, 105, 347
 autocatalitici di classe I e II, 357
 di gruppo I, 357
 di gruppo II, 357
 ipotesi del mondo a RNA, 471
 IPTG (isopropiltiogalattoside), 255
 IRE (*Iron Responsive Element*), 428
 IRES (*Internal Ribosome Entry Site*), 410
 IS (*Insertion Sequence*), 101, 213
 ISE (*Intronic Splicing Enhancer*), 366
 isocore, 103
 isola di patogenicità, 99
 isolatori, 307
 isole CpG, 296, 452
 isoschizomeri, 518
 ISS (Intronic Splicing Silencer), 366
 istoni, 84
 dimero, H2A-H2B 87, 172
 histone-fold, 86-87
 ottamero H2A, H2B, H3, H4, 85
 tetramero H3-H4, 87, 172
 varianti, 95, 460
 ISWI (complesso di rimodellamento della cromatina), 459

K

knock-down di un gene, 574

L

λ gt11 529, 545
 λ integrasi, 209
lac operone

lacA, gene, 254
lacI, gene, 254, 259
lacY, gene, 253
lacZ, gene, 253

lagging strand, nella replicazione del DNA, 154
 leading strand, nella replicazione del DNA, 154
 legame fosfodiesterico, 15
 LexA, 189, 282
 librerie di cDNA, 539
 librerie genomiche, 536
 Lk (*linking number*, numero di legame nel
 superavvolgimento), 34
 LINE (*Long Interspersed Nuclear Element*), 116, 118
 lisogenico, batterio, 275, 282
 lncRNA (*long non coding RNA*), 464, 468
 localizzazione degli mRNA, 440
 mRNA nei dendriti neuronali, 443
 mRNA ASH1 in lievito, 440
 mRNA nanos e bicoid in *Drosophila*, 442
 locus MAT (*Mating Type*), 220
 LTR (*Long Terminal Repeats*), 116, 213
 luciferasi, 570
 lunghi RNA non codificanti (lncRNA), 464

M

MALDI-TOF, 584
 mappa fisica, 591
 mappa genetica, 591
 MAR (*Matrix Attachment Region*), 76
 Maskin, 429
 MAT (*Mating Type*) 220, 221
 matrice nucleare, 75, 89
 maturazione dell'rRNA, vedi processamento dell'rRNA
 MBD proteine (*Metil-CpG-Binding Domain protein*), 454
 MCS (*Multiple Cloning Site*), 522
 mediatore, 299
 metagenomica, 565
 metil transferasi, 452, 457
 metilasi, 452, 517
 de novo, 452
 di mantenimento, 452
 metilazione
 2'-O metilazione, 336
 del DNA, 449
 di proteine, 489
 delle CpG e trascrizione, 454
 7-metilguanosina (m⁷G), 339
 metiloma, studio del, 602
 MGE (*Mobile Genetic Elements*), o trasposoni, 101
 microarray, 505
 micronucleo e macronucleo dei protozoi ciliati, 80, 183
 microRNA (miRNA), 464
 Lin-4, 464

microsatelliti, 113
 migrazione della giunzione, 209
 mimetismo molecolare, 415
 minisatelliti, 113
 miRNA (microRNA), 335, 464
 miRNP, 466
 MMR (*MisMatch Repair*), 188, 195
 modello a bruco, 239
 modello a cassette (locus MAT), 221, 224
 modello ad accartocciamento (scrunch), 239
 modello animale, 608
 modello dei domini gemelli per la trascrizione, 245
 modello del passaggio transiente, 239
 modello del tetranucleotide, 6
 modificazioni lipidiche, 490
 modificazioni post-traduzionali degli istoni, 94, 457
 molteplicità di infezione, 285
 mondo a RNA, 359, 471
 morte cellulare programmata, 200
 motivo A-minore dell'RNA, 49
 mRNA procariotici ed eucariotici, 396
 mRNP (*Messenger RibonucleoProtein*), 355
 MS/MS, 585
 MTP (*Minimum Tiling Path*), 592
 MUSCLE (programma bioinformatico), 599
 mutazione dominante-negativa, 583
 mutazioni, 61, 133

- acquisizione di funzione (*gain of function*), 187
- Amber, 61
- condizionali, 173
- delezioni, 186
- di senso, 62
- frameshift, 186
- F, 256
- Inserzioni, 186
- lacI*⁻, 259
- lacI*^d, 259
- lacI*^s, 259
- letali-condizionali, 147, 148
- missenso, 186
- non senso, 62, 186
- O^c, 256
- Ochre, 62
- Opale, 62
- perdita di funzione (*loss of function*), 187
- puntiformi, 184
- sinonime o silenti, 185
- soppressori, 61
- temperatura-sensibili (ts), 148
- transizioni, 185
- trasversioni 185

 MutLα e MutLβ, 195
 MutSα e MutSβ, 195

N

NAT (N-acetil transferasi), 488
 NCBI (National Center for Biotechnology Information), 589, 604
 ncRNA, 461
 ncRNA (*non coding RNA*), 101, 464
 NER (*Nucleotide Excision Repair*), riparazione per escissione di nucleotidi, 191, 192

- GG-NER (*Global Genome NER*), 193
- TCR-NER (*Transcription-Coupled NER*), 193

 NGS (*Next Generation Sequencing*), piattaforme di sequenziamento di nuova generazione, 554
 NHEJ (*Non-Homologous End Joining*), 199, 565
 nick translation, 153
 NLS (*Nuclear Localization Signal*), 311, 376
 NMD (*Non-sense Mediated Decay*), 435
 NOR (organizzatore nucleolare), 122, 337
 Northern blot, 504, 566, 569
 NPC (*Nuclear Pore Complex*), 375
 NSMD (*Non-Stop Mediated Decay*), 439
 NTD, 264
 NtrC, 264
 nucleasi sequenza-specifiche, 565

- TALEN, 567
- ZFN, 567

 nucleotide batterico, 75
 nucleolo, 337
 nucleoporine, 376
 nucleosidi, 13
 nucleosoma, 84, 90, 92, 172

- nucleo (core) del nucleosoma, 86
- posizionamento estrinseco, 90
- posizionamento in fase (*phasing*), 90
- posizionamento intrinseco, 90
- posizionamento rotazionale, 90
- posizionamento traslazionale, 90

 nucleotidi, 12

O

OD (Densità Ottica), 498
 O-glicosilazione, 492
 oligo-capping, protocollo, 596
 operatore, 252, 253

- dominante in *cis*, 256

 operone, 252
 operone *lac*, 253

- controllo negativo, 253
- controllo positivo, 260
- regolazione globale, 253
- regolazione specifica, 253

 operone policistronico, 252
 operone *trp* (triptofano), 265
 ORF (*Open Reading Frame*), 60, 358, 439, 598

organismi modello, 608
 organizzatore nucleolare, *vedi* NOR
oriC, 138, 157
 origine della vita, 471
 origini di replicazione, 138
 di *Escherichia coli*, *vedi* *oriC*
 di lievito, *vedi* ARS
 di SV40, 165
 degli eucarioti metazoi, 140
 mappatura fisica delle origini, 146
 umane, 145
 origini precoci e tardive, 140, 144
 ormone insulina, 476
 orologio molecolare, 606
 ortologi, geni, 605
 ottamero istonico, 86
 OxyS, piccolo sRNA, 462

P

p21, 202, 487
 p32 (proteina P32 del fago T4), 426
 p53, 202
 PABP (*poli(A) Binding Protein*) 344, 409, 428
 PABPC1 (*poli(A) Binding Protein Cytoplasmic 1*), 345
 palindromiche, sequenze, 30, 518
 PAP (poli(A) polimerasi), 344, 346
 paradosso del valore C, 102
 paraloghi, geni, 605
 passaggio transiente, 239
 pattern discovery, programmi bioinformatici, 602
 PB (*Processing Bodies*), 444
 pBR322, plasmide, 519
 PCNA (*Proliferating Cell Nuclear Antigen*), 167, 192, 482
 PCR (*Polymerase Chain Reaction*), 547
 peptide leader, 268
 peptidil transferasi, 389, 398
 permeasi, 253
 PET-Seq (*Paired-End Tag sequencing*), 597
 piattaforma Illumina (per sequenziamento), 556
 piattaforme di sequenziamento di nuova generazione, 554
 PIC (*Pre-Initiation Complex*) della trascrizione, 297, 302
 pinza β (β -clamp), 161
 pirimidine, 13
 pirofosfatasi, 150
 PK (proteine chinasi), 483
 plasmide batterico Ti, 535
 PNPasi (polinucleotide fosforilasi), 345
 poli(U), 58
 polimorfismi, 114, 605
 politenia, 78
 politenizzazione, 79, 181

posizionamento dei nucleosomi, 90
 post-RC (*post-Replication Complex*), 179
 ppGpp (guanosina tetrafosfato), 424
 pppGpp (guanosina pentafosfato), 424
 pre-initiation complex, 177
 pre-RC (*pre-Replication Complex*), 177
 pre-miRNA, 466
 pre-spliceosoma, 353
 Pribnow box, *vedi* elemento -10
 primer extension, 569
 primer, *vedi* innesco
 pri-miRNA (miRNA primari), 465
 principio trasformante, 8
 probe (sequenza sonda), 502
 processamento dell'rRNA (RNA processing), 48, 330, 569
 processamento dei tRNA negli eucarioti, 332
 processamento dei tRNA nei procarioti, 332
 processamento dell'rRNA negli eucarioti, 329
 processamento dell'rRNA nei procarioti, 329
 profago, 210, 275
 profilo d'espressione genica, 506
 progetto "Genoma Umano", 590
 promotore, 240, 241, 245
 promotore minimo, 294
 promotore di Pol I, 291
 nucleo (core) del promotore, 291
 UPE (*Upstream Promoter Element*) o UCE, 291
 promotore di Pol II, 297
 CAAT box, 306
 DCE (*Downstream Core Element*), 296
 DPE (*Downstream Promoter Element*), 296
 enhancer, 296, 306
 GC box, 306
 GRE (*Glucocorticoid Responsive Element*), 306
 MRE (*Metal Responsive Element*), 306
 promotore minimo (core promoter), 296
 TATA box, 296
 UAS (*Upstream Activating Sequence*), 296, 304
 promotore di Pol III, 294
 box A, 294
 box B, 294
 proteasi, 476
 protein tag (o etichette), 525
 proteine ricombinanti, 522, 523
 proteoma, 475, 489
 proteosoma, 476, 481
 PRR (*Post-Replication Repair*), 195, 196
 pseudogeni
 duplicati, 110
 processati, 110
 pseudonodi, 49
 pseudouridilazione, 336
 PTC (*Premature Termination Codon*) 436

PTK (proteine tirosina chinasi), 484
 “puff” cromosomico, 80
 purine, 13

Q

quartetto di G, 29, 83
 quoziente di compattamento (o impacchettamento), 71

R

ramificazione, sito di, 352
 rappresentatività di una genoteca, 536
 rare cutter (enzima di restrizione), 517
 real-time PCR, 549
 RecA, 189, 282
 regione codificante (dell'mRNA), 396
 regola di Chargaff, 17
 replicazione
 dei telomeri, 169
 della cromatina, 172
 meccanismo nei procarioti, 147
 meccanismo negli eucarioti, 164
 semiconservativa, 134
 replicone, 136,
 lunghezza, 142
 numero, 142
 velocità, 142
 replisoma, 144, 148, 150
 repressione da catabolita, 261
 repressore di lambda, 276
 repressore Lac, 258
 alta affinità, 260
 bassa affinità, 260
 forma dimerica, 259, 278
 forma tetramerica, 260
 testa, dominio N-terminale, 258
 repressori, 251
 restriction point o R (checkpoint in cellule di
 mammifero), 175
 restrizione dell'ospite, 516
 retroregolazione, 286
 retrotrasposoni, 116, 213
 LTR, 213
 non-LTR, 213
 virali, *vedi* elementi LTR
 reverse editing, 369
 RF (*Release Factor*), terminazione della traduzione,
 416
 RF1, RF2, RF3, 416
 RF (forma replicativa di DNA fagico), 156
 rH1 (repressore di H1), 220
 riarrangiamenti genomici, 224

riassociazione del DNA, o rinaturazione, 499
 riassortimento degli esoni, *vedi* exon shuffling
 ribosio, 12
 ribosoma, 379
 centro della peptidil transferasi, 388
 centro di decodificazione, 388
 disassemblaggio, 383
 ricostituzione, 383
 sito A (amminoacilico), 387
 sito E (exit), 387
 sito P (peptidilico), 387
 struttura tridimensionale, 384
 riboswitch, 463
 ribozimi a testa di martello (hammerhead), 333, 461
 ricombinasi, 209
 in serina, 211
 in tirosina, 211
 ricombinazione
 genetica, 204
 non omologa o illegittima, 531
 omologa o generalizzata, 204, 205
 sito-specifica, 204, 209
 V(D)J, 224
 ricostituzione (del ribosoma), 383
 rimodellatore (rimodellamento della cromatina), 321
 rinaturazione del DNA, *vedi* riassociazione del DNA
 ring finger (dominio di proteine implicate nella
 ubiquitinazione), 479
 ripetizioni invertite, 30
 RISC (*RNA-Induced Silencing Complex*), 444, 467
 risolvasi, 209
 risposta SOS, 282
 risposta stringente, 423
 R-looping, tecnica, 347
 RNA
 primer, 160
 regolatori, 461
 interference, 573
 RNA M1, 461
 Tie Club, 43, 44
 struttura, 43
 RNA attenuato, 268
 RNA guida (RNAg), 372
 RNA guida (sgRNA), 569
 RNAi (interferenza da RNA), 573
 RNA polimerasi, 231
 del fago T7, 236
 struttura, 236
 Pol I, eucariotica, 288
 procariotica, 236
 Pol II, eucariotica, 288
 Pol III, eucariotica, 288
 RNA polimerasi IVa (Pol IVa), 278
 RNA polimerasi IVb (Pol IVb), 278

RNA polimerasi procariotica, 236

α -CTD, 237

fattore σ , 238, 240

oloenzima, 240

subunità α , β , ω , 237

subunità σ , 240

RNA-Seq, 564, 595

RNasi (ribonucleasi), 329

E, 329

III, 329

P, 332

Z (o tRNasi), 333

RNA TERC, 461

Roll, 25

rotazione controllata (meccanismo delle topoisomerasi), 40

rp (r-proteine, proteine ribosomali), 379, 381

RPA (*Replication Protein A*), 165, 201

rRNA (RNA ribosomale), 379

eucariotici, 28S - 18S - 5,8S, 330, 337

procariotici, 23S - 16S - 5S, 329

rRNA, gene eucariotico, struttura

ETS, 331

ITS, 330

NTS, 329

RT-PCR, 506, 548, 574

Rut (utilizzo di Rho), 248, 249

S

Saccharomyces cerevisiae, 80, 440, 610

SAR (*Scaffold Associated Region*), 76

satellite "criptico", 112

satelliti α , *vedi* alphoid DNA

scansione (della subunità 40S lungo la 5'UTR), 405

screening

con sonda di acidi nucleici, 542

funzionale, 541

mediante anticorpi, 545

SDSA (*Synthesis-Dependent Strand Annealing*), 208, 209

seed sequence, 467

seed region, 467

segmento G (invertibile), del fago Mu, 218

segnale

di cleavage o taglio, 342

di poliadenilazione, 342

SELEX (*Systematic Evolution of Ligands by Exponential enrichment*), 576

sensibilità alle nucleasi, 93

sequence logo, 600

sequenza/e

a monte (downstream), 232

a valle (upstream), 232

consenso, 240, 591

d'inserzione o elementi IS, 213

di espansione dell'rRNA, 380

di Kozak, 408

di Shine-Dalgarno, 401

EST (*Expressed Sequence Tag*), 595

leader, 268

palindromica, 30

ripetute, 110

segnale, 476

sonda (probe), 502

uniche, 111

sequenziamento del DNA, 554

metodo chimico di (Maxam e Gilbert), 550

metodo enzimatico (Sanger), 552

sistemi di terza generazione, 560

sequenziamento dell'esoma, 564

serina/treonina chinasi, 484

SG (*Stress Granules*), 444

She1, She2, She3, 441

shuttle vector (vettori navetta), 533

silencer (silenzianti), 307

esonici, 366

intronici, 366

SINE (*Short Interspersed Nuclear Element*), 118

sintenia, 629

sintetasi, *vedi* amminoacil-tRNA sintetasi

siRNA (*small interfering RNA*), 573

Sistema CRISPR/Cas9, 568

sistema SOLiD (per sequenziamento), 557

sistema SOS, 189

sistemi di sequenziamento di terza generazione, 560

Pacific Bioscience, 561

Oxford Nanopore, 562

sistemi di sintesi proteica *in vitro*, 418

sito/i

accettore di splicing, 351

cos, 74

d'inizio della trascrizione (TSS), 232

di poliadenilazione, 342

di ricombinazione, 212

di taglio (degli ER), 517

di utilizzo di Rho (rut), 248

donatore di splicing, 351

ipersensibili, 93, 94

SL RNA (*Spliced Leader RNA*), 360

Sm, proteine, 352

SMC (*Structural Maintenance of Chromosome*),
proteine, 89

sncRNA (*small non coding RNA*), 100

snoRNA (*small nucleolar RNA*), 332, 338, 461

C/D snoRNA, 338

H/ACA snoRNA, 338

SNP (*Single Nucleotide Polymorphism*), 605

- snRNA (*small nuclear RNA*), 352
 U1, U2, U4, U5 e U6, 352
 snRNP (*small nuclear RNP*), 352
 solenoide (o fibra cromatinica da 30 nm), 87
 soppressori Amber, 62
 soppressori extragenici, 582
 sostituzioni non-sinonime, 606
 sostituzioni sinonime, 606
 Southern blot, 503, 544
 SPB (*Spindle Pole Body*), 173
 specificazione (nello sviluppo di *Drosophila*), 619
 spettro di assorbimento, 497
 spettrometria di massa, 584
 spliceosoma, 349, 352
 minore, o U12, 353
 splicing, 347
 alternativo, 363
 dei tRNA, 350
 di introni di gruppo I, 357
 di introni di gruppo II, 357
 inverso, 359
 nucleare, 349
 sporificazione in lievito, 611
 SPR (*Surface Plasmon Resonance*), 579, 586
 sRNA, 462
 snRNA, 294, 461
 oxyS, 462
 SSB (*Single Strand DNA Binding protein*), 158
 SSR (*Simple Sequence Repeats*), 115
 SsrA, *vedi* tmRNA
 stacking, *vedi* impilamento
 START (checkpoint in lievito), 174
 Staufen, 443
 sticky end, estremità coesive, 518
 STR (*Short Tandem Repeats*), *vedi* SSR
 strand passage (passaggio del filamento, meccanismo delle topoisomerasi), 40
Streptococcus pneumoniae, 6
 struttura modulare dei promotori, 304
 struttura modulare dei trans-attivatori, 309
 subunità ribosomali, 380, 384
 SUMO (*Small Ubiquitin like Modifier*), 481
 sumoilazione di proteine, 481
 superavvolgimento del DNA, 33
 numero di legame, 34
 twist, 34
 writhe, 34
 SWI/SNF (complesso di rimodellamento), 458
- T**
- TAF (*TBP Associated Factors*), 292
 T-Ag, *vedi* antigene T
 T-loop (*Telomere loop*), 83
 TAP (*Tandem Affinity Purification*), 527
 TAP-tagging, 527
 TATA-less, promotori, 292
 TBP (*TATA binding protein*), 292
 TBP (*Ter Binding Protein*) o Tus, 162
 TCR-NER (*Transcription-Coupled Repair NER*), 193
 T-DNA, 535
 tecnica di microiniezione in oociti, 625
 tecnologia 454 (per sequenziamento), 555
 telomerasi, 169, 217
 telomeri, 81-84
 template switch, 196
 teoria del tentennamento, *vedi* vacillamento
ter (terminatore della replicazione), 162
 TERC (*Telomerase RNA Component*), 170
 terminazione
 della replicazione, 162
 della traduzione nei procarioti, 416
 della trascrizione, 268, 344
 TERT (*Telomerase Reverse Transcriptase*), 170
 test del DNA, 114
 tetraloop, 49
 TH (*Two Hybrid*), o doppio ibrido, 587
 Tilt, 25
 tirosina chinasi, 484
 T-loop (*Telomer loop*), 83
 TLS (sintesi di DNA translesione), 168, 196
 TLS polimerasi, 197
 Tm (*melting temperature* = temperatura di fusione), 498
 tmRNA, 439
 topi
 knock-in, 631
 knock-out, 631
Mus musculus, 628
 topoisomerasi, 34, 39
 passaggio del filamento, 40
 rotazione controllata, 40
 topoisomerasi II, 89
 topoisomeri, 36
 tracrRNA, 569
 trans-esterificazione, 351, 357
 trans-splicing, 360
 intergenico, 362
 intragenico, 362
 trascrittasi inversa (*reverse transcriptase*), 213, 539
 trascrittoma, 564, 593
 trascrizione e rimodellamento della cromatina, 456
 trascrizione, regolazione negativa e positiva, 251
 trascrizione, terminazione, 246
 terminatori intrinseci, 246
 terminatori Rho-dipendenti, 248
 trascrizione, velocità nei procarioti e negli eucarioti, 233

trasduzione, meccanismo di scambio genetico, 99
 trasferimento di DNA nei vegetali, 535
 trasfezione, 530
 stabile, 531
 transiente, 530
 trasformazione, trasferimento di DNA all'interno di cellule, 99, 530
 traslocazione (del DNA nella testa del fago), 73
 traslocazione (nell'allungamento della traduzione), 410
 trasposasi, 213
 trasposizione del DNA, 204
 trasposoni, 212
 a DNA, 116
 autonomi, 213
 composti (o compositi), 213
 copia di *Drosophila*, 118
 LTR, 116
 non autonomi, 213
 non-LTR, *vedi* LINE
 Ty di lievito, 118
 trasversioni, 185
 TRCF (*Transcription-Repair Coupling Factor*), 244
 trigger loop, 302
 tRNA (RNA transfer, RNA di trasferimento), 389
 braccio accettore, 390
 braccio, D 390
 braccio dell'anticodone, 390
 braccio TΨC, 390
 braccio variabile, 391
 isoaccettori, 392
 struttura a L, 51
 struttura a trifoglio, 51
 tRNA_i^{Met} (tRNA di inizio eucariotico), 405
 TSS (*Transcription Start Site*), 232, 289
 twin domain model, 245
 Twist (nel superavvolgimento), 25, 34

U

U12, *vedi* spliceosoma minore
 UBC (*Ubiquitin Conjugating Enzyme*) o E2, 479
 ubiquitin ligase o E3, 479
 ubiquitina, 478
 ubiquitinazione di proteine, 478
 Ubl (*Ubiquitin-like proteins*), 481
 UCSC (University of California, Santa Cruz), 604
 unità di trascrizione, 232
 uORF (upstream ORF), 430
 UPE (*Upstream Promoter Element*), 291

Upf1 e Upf2 (proteine coinvolte nel NMD), 437
 UTR (*UnTranslated Region*, regione non tradotta), 397
 3'UTR (o sequenza trailer) dell'mRNA, 397
 5'UTR (o sequenza leader) dell'mRNA, 397

V

vacillamento, 64
 varianti istoniche, 460
 variazione antigenica, 223
 variazione di fase, in *Salmonella*, 219
 velocità e accuratezza della traduzione, 419
 vettore di clonaggio pUC19, 522
 vettori (lambda) d'inserzione, 528
 vettori (lambda) di sostituzione, 528
 vettori d'espressione, 523
 vettori di clonaggio, 519
 vettori di lievito, 531
 vettori navetta, *vedi* shuttle vector
 vettori per cellule animali, 534
 vita media degli mRNA, 395, 432
 VSG (*Variable Surface Glycoprotein*), 223
 VSG ELC (copia telomerica attiva delle VSG), 224
 VSG geni di base, 223

W

WES (*Whole Exome Sequencing*), 564
 Western blot, 525, 566, 569
 WGS (*Whole Genome shotgun Sequencing*), 592
 writhe, nel superavvolgimento, 34

X

Xenopus laevis, 140, 622
 Xis (escissionasi), 285
 XRCC1, 192

Y

YAC (*Yeast Artificial Chromosomes*, di lievito), 533
 YCp, 531
 YIp, 531
 YRp, 531

Z

zip-code, sequenze di localizzazione di mRNA, 440